University of Nevada, Reno

The Hazard of Graduation: Analysis of Three Multivariate Statistics Used To Study Multi-institutional Attendance

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Educational Leadership

by

Jessica Marie Muehlberg

Dr. Patricia Miltenberger/Dissertation Advisor

May 2013

We recommend that the dissertation
prepared under our supervision by

JESSICA MARIE MUEHLBERG

Entitled

The Hazard of Graduation: Analysis of Three Multivariate Statistics Used To Study Multi-Institutional Attendance

Be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

Patricia Miltenberger, Ed.D., Advisor

Rita Laden, Ed.D., Committee Member

Bill Thornton, Ph.D., Committee Member

Melisa Choroszy, Ph.D., Committee Member

Paul Neill, Ph.D., Graduate School Representative

Marsha H. Read, Ph.D., Dean, Graduate School

May 2013

Abstract

Adelman (2006) observed that a large quantity of research on retention is "institution-specific or use institutional characteristics as independent variables" (p. 81). However, he observed that over 60% of the students he studied attended multiple institutions making the calculation of institutional effects highly problematic. He argued that the student, and not the institution, should be the unit of analysis when studying the societal impacts of postsecondary education. Ranco (1996) predicted that our current measures of success (4-year and 6-year graduation rates), as well as persistence rates (the measure of the number of students returning year-to-year) were soon to become outdated. All of these measures are dependent on a student remaining continuously enrolled at one single institution. Whatever the term used, current methods of measuring student success may become inadequate to explain the realities of college student attendance and persistence. College attendance patterns are no longer adequately represented using a linear model (Sturtz, 2006). Statistics capable of using multi-institutional attendance data need to be explored to determine which are not only capable of evaluating these complex data sets, but which are simple to understand and use so that policy makers and administrators may take ready advantage of them. This study reviews the common methodologies for studying student enrollment patterns and examines novel methodologies that may improve the analysis of multi-institutional attendance. This also study explored how a thorough understanding of several statistics (logistic regression, discriminant analysis, and survival analysis) can be applied to the study of multi-institutional attendance and how researchers and administrators can best select which tool should be used to understand the impacts of student completion.

Dedication

I am dedicating this dissertation to my family, most especially…

To Lauren, Taylor, and Amelia for reminding me the wonder of learning;

To Lacey for being my rock;

To Dan for constantly reminding me not to take myself too seriously;

To Dad for keeping me centered; and

To Mom for always knowing I could.

This dissertation could not have been completed without your unyielding love and unending patience. Yes, you can now say that I am done.

Acknowledgments

Thank you to my committee whose support and guidance enabled this project to be completed. In particular, thank you to my advisor, Patricia Miltenberger who patiently guided me from wanting to study the world down to something manageable and achievable. Thank you also to Livia D'Andrea, who agreed to spend time investigating novel statistics…just because.

Thank you to my writing support group, Rich, Matt and Heather, without whose constant support, and Facebook blocking programs, I would not have survived. Those Saturdays turned out to be well spent.

Last but not least, thank you to my mentor, Melisa Choroszy, without whose constant prodding I would have ever embarked on this journey. You believed I could, when I wasn't too sure. I constantly learn from you every day.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

> Performance, even in relatively straightforward terms like graduation and retention, eludes our attempts to measure it. Students no longer march lockstep through four years of college to graduation. The educational model of the new millennium will likely be characterized more by lifelong learning, where students move among various kinds of higher education institutions, stopping in and out as their lifestyles and educational needs dictate. Enrollment patterns like these make indicators like four- or even six-year graduation rates for first-time, full-time freshmen largely irrelevant. Transfer students deprive both the sending and receiving institution of retention and graduation credit. (Ranco, 1996, p. 1)

As early as 1990, De los Santos and Wright observed that student enrollment pathways through our institutions were becoming increasingly complicated in patterns which they referred to as ‘swirl’. During the study of three national cohorts, with data spanning nearly thirty years, Adelman (1999-2006) observed an increase in the number of students attending multiple institutions. In 1999, Adelman reported that over half of all college students would attend more than one institution during their college career and shortly thereafter in 2006 he reported that figure increased to over 60%. However, not all sources show such high numbers. In a 2011 report published by the National Student Clearinghouse Research Center, no more than 8% of students nationally were enrolled in more than one institution during the 2010 – 2011 school year, with the majority of these students attending both a public 4-year and a public 2-year institution during the observation period. Researchers at the National Student Clearinghouse also observed that during the 2010-2011 school year over three percent (3.2%) of the students were attending more than one institution simultaneously. Regardless of the percentage of students engaging in these pathways, enrollment pathways in general have become increasingly complex.

As Adelman argued in 2006, the interest in multi-institutional attendees is not necessarily in the numbers “increasing from 47 to 57 percent [since the 1970s]; it is more a question of *how* they attended more than one school, and in what combinations and order” (p. 62). The interest with this growing population is in understanding how these enrollment pathways affect student success. Research using national data sets indicates that these students are not as likely to obtain a baccalaureate. Adelman (2006) found that “among the major categories of students who attended more than one school…there is a broad range of bachelor’s degree completion rates” (p. 65) and that those students who frequently alternated between more than one institution were the least likely to graduate (Table 1.1).

Table 1.1

Enrollment patterns of 1992 12th-graders who attended more than one postsecondary institution and percentage earning baccalaureate degrees under each pathway

Institutional Combination	Percent of Total	Percent Earning Baccalaureate
Two or more 4-year institutions	31.8	82.3
Reverse Transfer (4-year to 2-year)	9.7	<0.1
Classic Transfer (2-year to 4-year)	24.0	58.1
Swirlers (Alternating enrollment)	15.4	39.1
4-years with incidental 2-year work	14.3	86.4
Other	4.7	4.6

Note. Adapted from Adelman (2006).

Even as student enrollment pathways become increasingly more complex and include attendance at multiple institutions, methods for measuring institutional effectiveness remain the same (Sturtz, 2006). Sturtz stated, “most institutional effectiveness models are rooted in linear formulas used to measure graduation and retention rates for first-time, full-time, matriculated undergraduates who enter in the fall semester” (p. 151). He argued that this “growth of multi-institutional attendance and discontinuous enrollment" makes it challenging to meet the assumptions required to perform common statistical analysis and that these students are frequently excluded in the populations used to determine an institution's effectiveness. McCormick (2003) called for the need for more sophisticated measurement tools for multi-institutional attendance patterns. He advocated that "we need to develop a more sophisticated understanding of the various ways that students combine enrollment at multiple institutions" (p. 22) and that we need to understand how multi-institutional attendance affects educational outcomes. Borden (2004) illustrated this difference when comparing student tracking between linear models and multi-institutional attendance models. In linear models tracking is accomplished through measuring "first-time, full-time retention, and graduation rates within an institution" (p. 14). However, in models based on multi-institutional attendance student tracking becomes more complex, requiring "tracking within and across institutions, monitoring student progress according to mode of entry, collaborative tracking with high school and transfer feeds, and tracking transfers within and across state lines" (p. 14).

These changes in enrollment pathways and the need to subsequently change ways of measuring student success will also have more far reaching effects. As McCormick (2003) pointed out, "we will need far more nuanced ways to think about student careers and to categorize these complex attendance patterns. We may also need to revise policies and practices in light of these patterns" (p. 14). He argued that multi-institutional attendance has policy implications affecting "institutional finances, student financial aid, movements to promote institutional assessment and accountability, research on college impact, student advising, student assessment, and curriculum planning" (p. 23). With these wide-reaching effects, it is important to find the right tools to measure the impacts of multi-institutional attendance.

Statement of the Problem

Adelman (2006) observed that a large quantity of research on retention is “institution-specific or use institutional characteristics as independent variables” (p. 81). However, he observed that over 60% of the students he studied attended multiple institutions making the calculation of institutional effects highly problematic. He argued that the student, and not the institution, should be the unit of analysis when studying the societal impacts of postsecondary education. Etwell (2004) also stressed the importance of student based indicators versus institution based indicators.

Ranco (1996) predicted that our current measures of success (4-year and 6-year graduation rates), as well as persistence rates (the measure of the number of students returning year-to-year), were soon to become outdated. All of these measures are dependent on a student remaining continuously enrolled at one single institution. Whatever the term used, current methods of measuring student success may become inadequate to explain the realities of college student attendance and persistence. College attendance patterns are no longer adequately represented using a linear model (Sturtz, 2006). Statistics capable of using multi-institutional attendance data need to be explored to determine which are not only capable of evaluating these complex data sets, but which are simple to understand and use so that policy makers and administrators may take ready advantage of them.

Purpose of the Study

The purpose of this study was to review the current methodologies for studying student enrollment patterns and to examine novel methodologies that may make better use of complex data sets. In addition to reviewing how these pathways are studied, this study also reviewed the most recent research describing the classification of these various enrollment pathways and how these pathways are represented in student unit level enrollment data. This study ultimately explored alternative methodologies to study these more complex data sets, in order to advise researchers and administrators on what tools could be used to understand the impacts of this phenomenon on student completion.

Research Question

The data set included a random sample of 600 students who graduated high school in 2000 and were subsequently admitted into a baccalaureate granting program at a mid-sized land-grant institution in the western United States. Enrollment records from all institutions attended were combined with a limited number of variables from the primary institution's student information system. The primary question was to what extent are the results of three statistical methodologies (logistic regression, discriminant analysis, and survival analysis) comparable? How does the selection of specific statistics affect the sample selection, periods of observation, assumptions, variables used and outcomes (what questions can be answered) in a study of multi-institutional attendance? Lastly, how do administrators ensure that they are evaluating and correctly applying persistence and retention research to their respective institutions?

Definition of Terms

Several terms are used throughout the study. While the evolution of these terms is explored fully in Chapter Two, a succinct definition of each is included below.

"Enrollment Pathway" An enrollment pathway is defined by the types and sequence of term enrollments initiated during a student's postsecondary career.

"First-time Full-time Cohort" Cohort of students most often measured in higher education, defined as all of the new freshmen enrolled full-time at a specific institution. This cohort serves as the base of most institutional retention and graduation metrics.

"Institutional Retention" Metric defined as the percentage of a specific group of students enrolling in a subsequent term after being enrolled in a base term.

"Multi-institutional Attendance" The act of enrolling in more than one institution at the same time.

"Student Persistence" Metric defined by the student enrolling in a subsequent term after being enrolled in a base term.

"Transfer" The act of enrolling at a different institution than previously enrolled.

Framework of Institutional Retention

The most frequently used measure of success and performance in higher education is the graduation rate (Archibald & Feldman, 2008; Astin, 2005-2006; Cohen & Ibrahim, 2008; Jones-White, Radcliffe, Huesman, & Kellogg, 2009; Russell, 2009). The federal government requires public institutions of higher education to publish these data with the federal Student Right-to-Know Act of 1990 (SRTK). The SRTK act mandated the format of the graduation rates, defining the graduation rate as "first-time, full-time degree- or certificate-seeking students at 150 percent of normal time" (Russell, 2009, p. 2). The same cohorts defined by the SRTK are also used by institutions to measure retention. This framework is explored for its sufficiency to measure the effects of multi-institutional attendance.

Methodology

This is a qualitative study comparing three different quantitative statistical methods. The study utilized one single data set, primarily term based enrollment data, including 600 students who graduated from high school in 2000 and had at some point enrolled as a degree seeking baccalaureate student at a mid-sized land-grant institution in the western United States. Observations spanned a period of ten years from Fall 2000 through Spring 2010.

For each statistic used (logistic regression, discriminant analysis, and survival analysis) the dataset was transformed to meet the specific assumptions required for each statistic as described in Chapter 3. Each statistic provided insight about multi-institutional attendance. The goal of this study was to compare the use of each. This study evaluated issues such as assumptions needing to be met, which variables may or may not be excluded, and limitations on the data that can be included within the analysis.

The framework for this evaluation used three criteria. The first was usefulness of the individual statistic in answering questions related to multi-institutional attendance. The second was ease of use of the statistic and the interpretation of its results. In order to provide useful answers to policy makers and administrators, the results and implications of the individual analysis must be clear. Third was the ability of the statistic to use as many of the cases and variables without significant transformations or exclusions as possible. Unique enrollment pathways, and cases that often vary from the norm, characterize multi-institutional attendance. In order to perform the most robust analysis, it is ideal to include as many cases as possible.

Limitations

The primary limitation in this study is that while the random sample of 600 students was selected from all of the high school students that graduated in 2000, it was limited to just those students who applied and eventually attended the primary institution used in this study. While this sample gave a well-defined picture of enrollment pathways, there could be systematic differences with those who chose to attend higher education elsewhere. Further studies should try to include all students who graduate high school at the same time regardless of which institution they enrolled.

Lastly, the variables used in each of the statistical analyses were chosen based on the availability and commonality in other enrollment based studies. There may be other variables that could have been included to provide a more robust snapshot of how multi-institutional attendance affects student outcomes such has when English and math requirements were completed, average term grade point averages, number of withdrawals. However, the goal of this study is to compare the statistics themselves and not the direct effects of multi-institutional attendance.

Delimitations

The primary delimitation in this study is the use of a regional data set. While enrollment data was obtained from multiple institutions, each student included in this study will have attended the University of Nevada, Reno for at least one semester as a degree seeking undergraduate student. While there are large national enrollment datasets available through the National Center for Education Statistics, local conditions are also important in the study of higher education. The goal of this study is more about establishing the framework for how the phenomenon should be explored and not on the phenomenon itself.

A secondary delimitation is the selection of just three statistics to compare. While the selection of logistic regression, discriminant analysis, and survival analysis yielded information as to how the selection of a specific statistic can affect the results of a study, there are many other statistics that may also be appropriate for this data set. Statistics such as cluster analysis may reveal information as to whether those students who graduate belong to a specific subgroup with its own characteristics and traits.

Organization of this Study

This study is organized into five parts. The first part is the introduction, which defined the context and reasons for why this study is important and defined the terms used throughout the study. The second chapter reviews the literature contributing to the current understanding of both enrollment pathways and the statistics that have been employed to measure them. The third chapter of this study identifies the methodology applied in the study and describes how this study was conducted. The fourth chapter describes the results of each statistic employed. Finally, the fifth chapter discusses the impacts and implications of the results obtained and how researchers and administrators could approach similar studies.

Chapter Two: Literature Review

Hagedorn (2006) observed that “according to the Tinto theory, university students must abandon their old lives and become socially and academically integrated into their new postsecondary environments in order to succeed” (p. 9). She argued that community college students do not abandon their old lives and that this phenomenon may apply to university students as well. The stereotypical view of the college pathway is that high school graduates go “do the college thing” for four to six years and then graduate. While this view has been slowly evolving, retention research is still rooted in the work of Astin (1993), Pascarella and Terenzini (2005), Tinto (1993, 2006-2007) and others who base their models upon the premise that a key component for retention is the interactions between the student and their respective college. In recent years, researchers have observed that students are increasingly attending more than one college, and to complicate the matter further, they are increasingly likely to attend more than one college *at the same time*. Adelman (1999) described the phenomena as “students [are] filling their undergraduate portfolios with courses and credentials from a variety of sources, much as we fill our shopping bags at the local mall” (p. 39).

Researchers have begun to explore the different pathways that students take on their way to graduation as there is increasing accountability using performance metrics such as graduation rates, persistence and retention (Shapiro et al., 2013). In the past twenty years, researchers have proposed numerous terminologies to describe these new pathways, including “swirl” (de los Santos & Wright, 1990), “multi-institutional attendance” (McCormack, 2003), “pipelines” (Adelman, 1999), and “enrollment/attendance pathways” (Adelman, 2006a). While researchers are just starting to tackle the complexity of this phenomenon, the concept is not new. Researchers have been studying one aspect of this phenomenon for decades in the form of transfer (Townsend, 2001).

What is Success in Higher Education?

Describing success: retention, persistence, and engagement. Adelman (2006a) eloquently defined retention and persistence; "institutions 'retain'; students 'persist'" (p. 107). He argued that the study of retention focuses, in essence, on how well institutions hold on to their students. However, the study of persistence transfers the focus to the student and their movement through postsecondary education. "In the rhetoric of 'retention,' students are passive: Something is done to them, and that 'something' assumes a deficit model. Under the rhetoric of 'persistence' they are actors shaping their fate, with a model of success in mind" (Adelman, 2006a, p. 107). Reason (2009) distinguished the manner in which retention and persistence are measured. Retention is typically measured as a percentage of students from a specific cohort who re-enroll during a specific observation period. Persistence is harder to measure, as it is dependent upon the student's ultimate goal, which may, or may not, be a degree. Reason argued that "because individual students define their goals, a student may successfully persist without being retained to graduation" (p. 660). Regardless, however, the most commonly measured outcome for both retention and persistence is graduation rate, utilizing the federal standard of measuring the percentage of students who graduate within four years or six years.

One of the most prominent findings of retention and persistence research as it relates to multi-institutional attendance is that institutions matter, namely that the “students’ interactions with their environments matter” (Reason, 2009, p. 675). In Tinto’s model of student integration, he makes explicit connections between the academic and social systems, or environment, of an institution to the outcome of student retention (Tinto, 1993, 2006-2007; Wolf-Wendel, Ward, & Kinzie, 2009). Pascarella and Terenzini (2005) found that where a student chose to attend college had less of an effect on measured outcomes of success and development than the student’s individual experiences. However, they did note that this data is complicated by the recent observations that “more than half of students who initially enroll at a four-year college ultimately attend two undergraduate institutions, and nearly 40 percent attend three or more” (p. 591).

Kuh (2005, 2009) defined engagement as “the time and effort students devote to activities that are empirically linked to desired outcomes of college *and* what institutions do to induce students to participate in these activities” (p. 683). Kuh (2009) argued that engagement is a two-way street in that both institutions and students have responsibilities for determining student success. He argued that institutional policies, espoused beliefs, and how institutions organized learning experiences have direct impacts on a student’s success (Kuh et al., 2005; Kuh, 2009). Pascarella and Terenzini (2005) empirically demonstrated that student engagement is positively associated with student outcomes such as graduation.

The challenge to the generalizability of current research on retention, persistence, and engagement is that most researchers tend to exclude adults, non-residential campuses, two-year institutions, and students who attend more than one institution (Pascarella & Terenzini, 1991; Metz, 2004; Wolf-Wendel, Ward, & Kinzie, 2009). Metz

(2004) found that researchers "struggled to identify causal models of student persistence and integrate generally accepted theoretical constructs on student persistence, grounded in the four-year college environment, into the two-year college sector" (p. 198). The literature is lacking in the investigation of how attendance at multiple institutions, both sequentially and simultaneously, affect the persistence and retention of multi-institutional attendees. The importance of the institution is well established, but the interactional effects of multiple institutions are not.

Measuring success: graduation rates. The most frequently used measure of success and performance in higher education is the graduation rate (Archibald & Feldman, 2008; Astin, 2005-2006; Cohen & Ibrahim, 2008; Jones-White, Radcliffe, Huesman, & Kellogg, 2009; Russell, 2009). The federal government requires public institutions of higher education to publish these data with the federal Student Right-to-Know Act of 1990 (SRTK). The SRTK act mandated the format of the graduation rates, defining the graduation rate as "first-time, full-time degree- or certificate-seeking students at 150 percent of normal time" (Russell, 2009, p. 2).

While these measures have become universal, their use is controversial. For example, Archibald and Feldman (2008) argued that institutions should not aim for a 100% graduation rate as it is not always in the best interest of the student to graduate. They point out that many students' learning objective includes gaining new skills, which may or may not coincide with obtaining a certificate or degree. Additionally they argued that institutions can increase graduation rates by using unsound policies, such as increasing admissions standards to only admit those students who have the best chance of success or by “lowering curricular standards or by encouraging more grade inflation” (p. 81).

Other challenges to using raw graduation rates are articulated by Roska (2009), Linn (2009), Russell (2009) and Adelman (2006b). Roska (2009) found that graduation rates at public four-year institutions “may be artificially inflated or deflated depending on the presence of community colleges” (p. 15). Linn (2009) argued that the current measurements exclude too many students, ignore other important student milestones, and that the data are not disaggregated enough to be useful in identifying possible interventions. Russell (2009) also challenged the use of the current graduation rate metric (GSR) as it is:

> …based on an outmoded model of student behavior that assumes linear and timely progression through a single institution. This model fails to recognize the increasingly common “swirling” behavior that involves alternating full- and part-time attendance, enrollment in multiple institutions, transfer, and stopping out. Specifically, the GRS cohort (the “denominator”) excludes part-time students, adults with prior college coursework, students who began elsewhere but transferred into a particular institution, and students who began in any term other than fall. The GRS definition of “success” (the “numerator”) excludes students who transfer and graduate elsewhere and students who take longer that [sic] the allotted time to graduate. By not taking into account students’ actual enrollment behavior, and by failing to encompass much of what institutions do, current graduate rate data lead to misleading conclusions about institutional performance. (p. 4)

Adelman (2006b) noted, based on similar observations, that “roughly half of traditional-age undergraduates are excluded from the Education Department’s calculation of graduation rates.” Jones-White et al. (2009) observed that there is a need to “broaden the definition of student success to include degree completion beyond the originating school and expand understanding of factors contributing to a more expansive definition of student success” (p. 155). They point out that the new Voluntary System of

Accountability uses multi-institutional retention and graduation rates as a method to describe differences in success among institutions of higher education.

Sturtz (2006) found that less than half of all new students in the Connecticut State University System over a period of three years (2002-2005) were actually included in the SRTK cohorts. He argued, “the progression of between 58 and 60 percent of new students is inconsequential in the accountability and policy arenas” (p. 153). He proposed that reporting should focus on the student and not the institution. Borden (2004) similarly found that at his campus nearly two-thirds of all baccalaureate recipients actually started their postsecondary career at another institution, thus removing them from being included in his institution’s graduation rates.

After reviewing data from 56,818 students, Astin (2005-2006) found that nearly two-thirds of the variation in graduation rates among institutions can be attributed to student characteristics, not institutional effect. Astin further argued that using SRTK graduation rates to evaluate institutions “discourage institutions from enrolling relatively poorly prepared students in order to maximize their raw retention rates” (p. 15). In a similar study, Scott, Bailey, and Kienzl (2006) found the difference in graduation rates between public and private institutions could largely be attributed to systematic differences in the respective student bodies. Furthermore, they found that even though public institutions have a lower overall graduation rate, they were just as effective, if not more so, in graduating their students as private institutions.

The SRTK cohorts are what drive the four- and six-year graduation rates that all baccalaureate degree-granting institutions are measured against. As multi-institutional attendance has increased, so has the difficulty in using data from these cohorts to measure student success. Once students move to another institution, even temporarily, they are no longer represented in any institution's reported graduation rate. Furthermore, researchers of higher education (Bahr, 2009; de los Santos & Wright, 1990; Li, 2010; McCormick, 1997) have used these cohorts as a sample of convenience to study and develop theories about student success and graduation. As these cohorts decrease as a proportion of the overall college population, the conclusions and interventions based on these groups become less and less generalizable.

The Study of Multi-Institutional Attendance

The study of multi-institution attendance is not a new phenomenon. Early studies of multi-institutional attendance focus on the transfer and reverse transfer phenomenon and date back to as early as the 1960's (Townsend, 2001). Studies of transfer students primarily focused on the migration of students from a community college to a four-year institution and back again. Starting in the 1990's researchers observed that student attendance patterns were far more complicated than previously thought, and the era of "swirl" began (de los Santos & Wright, 1990).

The national data sets used to study multi-institutional attendance. Of the 18 studies identified in this research that examine multi-institutional pathways, 11 use data from one or more of three national data sets (Appendix A). These three data sets were collected by the National Center for Education Statistics and include the National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HSB), and the National Education Longitudinal Study of 1988 (NELS:88).

NLS-72. Participants in the NLS-72 study were high school seniors in 1972. These data span 14 years and includes six surveys, high school records, and postsecondary transcripts collected in 1984. The last survey sample was selected from any participant who had responded to the initial survey and at least one additional survey. The total number of respondents to the last survey was 12,841 (Sebring et al., 1987).

HSB. The second was the High School and Beyond longitudinal survey which followed the sophomore class of 1980. The data for this longitudinal study spans 12 years and includes five surveys and a transcript study. The last survey included a sample of 14,825, while the transcript study included 9,064 members of the cohort that had reported postsecondary attendance (Zahs et al., 1995).

NELS:88. The last was the National Education Longitudinal Study of 1988 (NELS:88) which followed 12,144 eighth graders starting in 1988. The data in this study include five surveys over a period of 12 years, achievement tests, surveys of students' teachers, parents and school administrators, and lastly transcript data from both high school and postsecondary institutions (Curtin et al., 2002).

The transfer student. The most common of all multi-institutional attendees is the transfer student. Townsend (2001) noted that with the creation of the public junior college in 1901, the intent was that students would attend the first two years, obtain an associate degree, and then transfer to a four-year institution. The intent was that students would only transfer one direction and ideally only once. McCormick (1997) defined transfer as a "transition between postsecondary institutions in which the second institution (the destination or receiving institution) typically grants the student credit for coursework taken at the first institution (the origin, or sending institution)" (p. 1). He further argued that transfer students do not normally return to their first institution and students who have temporary enrollment at a second institution should not be considered transfer students.

In 1993, Barkley noted that fewer students obtained an associate degree prior to transfer and that an increasing proportion of transfer students were no longer limited to traditional students and “may well be anyone enrolled at the community college, be they in vocational or academic programs, young or old, recent high school graduates, or former high school dropouts” (p. 40). Grubb (1991) noted that students were increasingly transferring to four-year institutions prior to completing an associate degree. Using data from the National Longitudinal Study of the high school graduates from 1972 and the High School and Beyond Study of the high school graduates from 1980, Grubb found that 65% of the 1972 class transferred without earning an associate degree and 73% of the 1980 class did the same.

McCormack (1997), using the Beginning Postsecondary Students Longitudinal Study, tracked students from their first year (1989-90) in a postsecondary institution through 1993-94 and found that nearly half of the students had enrolled in at least one additional institution within the five-year observation period. However, only half of these students were considered transfer students. In addition to the typical transfer student, two subtypes were predominant in the literature: reverse transfer and lateral transfer.

Reverse transfer. In Townsend’s (2000) review of the literature about the reverse transfer phenomenon, she acknowledged that Clark first described the phenomenon in 1960 during a study of a California junior college. She defined reverse transfer students as those who have “already matriculated at, or even graduated from, a four-year college” (p. 302) prior to enrolling at a two-year institution. Researchers have identified two types of reverse transfer students: undergraduate or non-completers and post-baccalaureate or completers (Townsend & Dever, 1999; Townsend, 2000; Winter, Harris, & Ziegler, 2001). Undergraduate reverse transfer students differ from post-baccalaureate transfer students in that they had not obtained a baccalaureate degree prior to attending the two-year institution.

There are various published estimates of the number of reverse transfers. From a national survey administered to the deans of student personnel services, Heinze and Daniels (1970) found that of 46 community colleges responding, nearly 10% of community college students were reverse transfers. McCormack (1997) estimated that 13% of students who began college in 1989-90 eventually became a reverse transfer student. Adelman (1999) in reviewing a national sample of over 8,000 student transcript records estimated that 7.6% percent of undergraduates engaged in reverse transfer. In his subsequent work in 2006 that estimate had increased to 8.2%. Bach et al. (2000), in reviewing the transcript records of 334 Oregon students, found that 23% of the transfer students were reverse transfer students. Winter and Harris (1999) reviewed enrollments in Kentucky community colleges and found that reverse transfer students comprise approximately 11% of the two-year student population.

Lateral transfer. Lateral transfer is defined as the transfer of a student from one community college to another (Bahr, 2009; Peter & Cataldi, 2005) or transfer from one four-year institution to another (Kirk-Kuwaye & Kirk-Kuwaye, 2007; Li, 2010). Estimates indicate that between 13 and 28% of community college students are involved in at least one lateral transfer (Bahr, 2009; Peter & Cataldi, 2005). In Bahr's (2009) study of lateral transfer, he reviewed the educational records of 167,997 California community college students. He found that while over one-quarter of those students engaged in lateral transfer, very few of those engaged in serial lateral transfer or repeated lateral transfers during his observation period.

The various transfer student pathways defined above form the core of the phenomenon of multi-institutional attendance. Current estimates indicate they account for a large proportion of multi-institutional attendees. However, while these models were complex in their time, they no longer fully describe student attendance patterns.

The plethora of pathways. In the 1990's researchers began to observe that the movement of transfer students might be more complex than previously defined, with transfer students repeatedly moving from college to college, instead of being a single, discrete event (Adelman, 1999; Barkley, 1993; de los Santos & Wright, 1990; Sturtz, 2006; McCormack, 1997; Townsend, 2001). Adelman (2006a) pointed out that pipelines are "…unidirectional closed spaces, and under the 'pipeline' metaphor students are passive creatures…swept along or dropping out of the space completely through leaks at the joints" (p. 107). He further argued that pathways "…allow for multiple analyses and discoveries of tools that suggest (but do not predict) productive routes to education goals" (p. 108). The most common term applied to this pathway, or lack thereof, is swirl or multi-institutional attendance. The earliest use of the term "swirl" was by de los Santos and Wright in 1990. They observed that nearly one-third of students graduating with a baccalaureate from Arizona State University had earned credit from other institutions. Furthermore, they found that the pathways were not the clean, linear progression that typified the view of the time, but rather a complicated swirl between and among local two-year and four-year institutions.

Another challenge to the transfer model is the changing nomenclature. While transfer, reverse transfer, and lateral transfer have historically had clear definitions, when students started to swirl, the definitions became problematic. How does one classify a student who matriculates at a four-year institution, attends a two-year institution for a semester, and then returns to the institution of matriculation? While initially considered a transfer student in the most general sense of the word, the reality challenges the simplistic definitions. The first institution may consider the student a stop-out (someone who has a break in attendance), yet the second institution may classify him or her as a reverse transfer. McCormick (2003) summed up the problem succinctly when he stated that students with attendance at more than one institution could not be immediately categorized as transfer. In order to study the phenomena of multi-institutional attendance, the research must "cover a sufficiently long period to observe transfer behavior, allowing for part-time attendance, stop out, and eventual return to the first institution" (p. 19).

There have been two general approaches in the attempt to categorize and understand student multi-institutional attendance pathways. The first, a qualitative approach, is based on observation and anecdotal data (McCormack, 2003; Townsend, 2001). This type of approach tends to yield meaningful pathways to which researchers and policy makers can relate. These pathways however tend to be amorphous and are not always mutually exclusive. These characteristics make it difficult to employ the pathways in analyses as students may belong to more than one, or the classification of students into these pathways may not be well defined. Use of student intent for classification is hard for researchers to interpret when solely using transcript data. The second, a quantitative approach, relies solely on statistical methods to identify pathways (Adelman, 1999, 2000, 2003, 2004, 2005, 2006a; Bach et al., 2000; Goldrick-Rab, 2006; Marti, 2008). The strength of the quantitative approach is that all potential pathways are well defined and students can definitively be assigned to one pathway or another. The challenge with the quantitative approach is that it is often difficult to communicate what these pathways mean and the student motivation behind each pathway. This in turn provides a challenge to providing adequate interventions.

Qualitative pathways. De los Santos and Wright (1990) described three pathways, using previously defined categories of transfer and reverse transfer while adding a new category, that of concurrently enrolled students (those attending more than one institution at the same time). Peter and Cataldi (2005) used a similar classification scheme of transfer, reverse transfer and concurrently enrolled students in their study of multi-institutional pathways.

Townsend (2001) identified seven transfer patterns, categorized into two primary groups dependent upon whether a student starts at a two-year or a four-year institution. Students who started at a two-year institution can be classified in one of four groups:

- Transfer to a four-year institution prior to completing the two-year associate degree
- Transfer to a four-year institution with a transfer degree (Associate of Arts or Associate of Science)
- Transfer to a four-year institution with a nontransferable degree (Associate of Applied Science)
- Transfer from and to a community college repeatedly

Students starting at four-year institutions can be classified into one of three groups:

- Transferring high school dual credit courses offered by a two-year institution
- Transferring summer courses
- Transferring courses taken through concurrent enrollment

McCormick (2003) developed a schema to describe student multi-institutional attendance patterns that have been used as a baseline to study the multi-institutional attendance phenomena.

- One-way transfer: students complete a minimum of one semester and up to two years of course work at either a two- or four-year institution and then transfers to another institution
- Trial enrollment: students experiment with the possibility of transfer
- Special program enrollment: these include organized programs such as study abroad or semester at sea
- Supplemental enrollment: students enroll at another institution for one or two terms to supplement or accelerate their program, usually summer
- Rebounding enrollment: students alternate enrollment at two or more institutions
- Concurrent enrollment: student takes courses at two institutions simultaneously
- Consolidated enrollment: students satisfy the awarding institution's residency requirements, but a substantial share of their credits come from at least two other institutions
- Serial transfer: students make one or more intermediate transfers
- Independent enrollment: students pursue work unrelated to their degree program and no credits are transferred

McCormack noted that these definitions rely on anecdotal evidence and have yet to be operationalized. The challenge with operationalizing these categories is that not all are mutually exclusive. However the strength of these pathways is in the meaning that researchers can readily interpret from them.

Quantitative pathways. Bach et al. (2000) conducted a descriptive analysis of transfer student attendance patterns using 336 Oregon students who were undergraduates at their first university enrollment and who were not concurrently enrolled at a community college and university during the first semester of observation. Within their population, they identified 48 discrete attendance patterns. However six of those attendance patterns represented just over 80% of their sample (Table 2.1).

Table 2.1

Transfer Student Attendance Patterns

Pattern	n	%	Code	Classification
1-2	174	51.78%	1	Community College Attendance
2-1-2	39	11.91%	2	University Attendance
1-2-1	24	7.14%	3	Concurrent Attendance
1-2-1-2	15	4.46%		
1-2-3-2	8	2.38%		
2-1-2-1-2	7	2.08%		
1-2-1-3-2	5	1.49%		
All others (41)	64	19.04%		

Note. Adapted from Bach, S.K., Banks, M.T., Kinnick, M.K., Ricks, M.F., Stoering, J.M., and Walleri, R.D. (2000). Student attendance patterns and performance in an urban postsecondary environment. *Research in Higher Education, 41*(3), 315-330.

Goldrick-Rab (2006) argued that “multi-institutional attendance and discontinuous enrollment intersect in meaningful ways” (p. 66), defining four student pathways in her study of multi-institutional attendance. The four pathways defined by Goldrick-Rab were (1) traditional (continuous enrollment at a single institution), (2) interruption (discontinuous enrollment at a single institution), (3) fluid movement (continuous enrollment at more than one institution), and (4) interrupted movement (discontinuous enrollment at more than one institution). Goldrick-Rab analyzed the transcripts obtained from 4,628 students starting at a four-year institution as part of the National Education Longitudinal Study conducted by the National Center for Education. The goal of her study was not to necessarily describe enrollment pathways but to determine which students were most likely to follow the non-traditional pathways.

Marti (2008) used latent trajectory analysis to identify the enrollment pathways in his study of community college students. His data set, from the Florida Department of Education, included 3,490 students, representing 27 different Florida colleges. However, while he used statistical techniques to identify pathways, his technique clustered students into five qualitatively distinct patterns: full-time, long-term; two years and out; long-term decliners; part-time, long-term; and one term and out.

Adelman (1999, 2000, 2003, 2004, 2005, & 2006a) conducted the most prolific studies of multi-institutional attendance over the past decade. These studies used three national longitudinal data sets. The first was the National Longitudinal Study of the High School Class of 1972 (hereafter 1972 cohort). The second was the High School and Beyond Longitudinal Study of 1980 Sophomores (hereafter 1982 cohort). The last was the National Education Longitudinal Study of 1988 (hereafter 1992 cohort). Through all of these studies, Adelman described an increasing trend towards students attending more than one institution while an undergraduate (Table 2.2).

Table 2.2

Percent of the 1972, 1982, and 1992 cohorts who attended one, two, or more than two schools as undergraduates.

	Class of 1972 (1972-1984)	Class of 1982 (1982-1993)	Class of 1992 (1992-2000)
All who earned more than 10 credits			
One	52.5	48.7	43.5
Two	32.5	32.7	35.1
More than two	15.0	18.6	21.5
All who earned bachelor's degrees			
One	42.8	42.0	40.6
Two	38.2	36.7	36.7
More than two	19.0	21.3	22.7

Note. Adapted from Adelman, C. (2004). Principle indicators of student academic histories in postsecondary education, 1972-2000. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Adelman in 2003 described eight different mobility histories using the 1992 cohort data.

- Excursion (occasional change). In this group the student was based in a 4-year institution, earned 30 or more credits and either was an incidental community college student, participated in study abroad, or attended another institution(s) for summer terms only.
- Migration (permanent change). In this group the student either moved from 2-year to 4-year or vice versa without returning or earned an associate or baccalaureate degree from an institution other than where the student began.
- Fragmentation (extreme mobility). In this group the student earned at least 30 credits and either attended three or more institutions without earning a credential, had alternating attendance at both 2-year and 4-year institutions without earning a credential, or attended two or more institutions and was still enrolled at the end of the observation period and not a degree candidate.

- Discovery (extreme mobility with credential or momentum). This group mirrors that of the fragmentation group, except that the student earned a credential or was a degree candidate with a minimum of 60 earned credits.
- Limited participation. Students who have earned between 11 and 29 credits without a degree.
- Incidental students. Students who earned less than 10 credits.
- No mobility, but attending two schools. Student attends two institutions in sequence earning 30 or more credits.
- No mobility, one school. Student attends one institution while earning 30 or more credits.

Adelman (2004) estimated that since the 1970's 60 % of all baccalaureate degree recipients have attended more than one institution. Even within Adelman's work, similar but different enrollment patterns exist. Adelman identified ten categories using the attendance patterns of the 1982 and 1992 cohorts: (1) 4-year only; (2) 4-year, then 2-year; (3) 2-year, then 4-year; (4) alternating 2-year and 4-year; (5) 4-year plus incidental 2-year; (6) 4-year plus other sub-baccalaureate; (7) 2-year only; (8) 2-year plus other sub-baccalaureate; (9) other sub-baccalaureate only; (10) 4-year, 2-year, and other sub-baccalaureate.

Over the past twenty years, many researchers have defined complex pathways of student institutional attendance. As the pathways of students attending institutions of higher education become more complex, so must the means by which these pathways are studied. Researchers must adapt new methodologies that can account for this additional dimension of enrollment.

How Enrollment Pathways are Studied

Multiple researchers have investigated the effects of multi-institutional attendance on various aspects of student success, namely graduation. As illustrated in Appendix A, the studies of multi-institutional attendance have had varying methodologies. However, when analyzing multi-institutional attendance, there are four factors that contribute to the generalizability of a study: sample selection, period of observation, variables included, and statistical test used.

Sample selection. The study of multi-institutional attendance challenges current research standards in higher education. Researchers have noted that research in higher education has focused on samples consisting primarily of traditional-aged students, enrolling for full-time credit each term (Adelman, 2006a; Metz, 2004; Sturtz, 2006). This method of sample selection became codified with the introduction of the Student Right-to-Know Act (1996). This legislation mandated that institutions of higher education report graduation and retention data of new first-time, full-time freshmen. This allowed students and the public to compare institutions on a standard measure. A legacy of this model is that retention research has tended to select samples based on this cohort. As detailed earlier, the number of students in non-traditional attendance pathways has been increasing over the last two decades. Therefore, research based solely upon a first-time, full-time cohort is not as generalizable to these students.

Fortunately, many of the researchers who study multi-institutional attendance have also noted this challenge (Adelman, 1999, 2003, 2004, 2005, 2006a; Adelman, Tabs, & Berkovits, 2000; Alfonso, 2006; Bach et al., 2000; de los Santos & Wright, 1990; Golrick-Rab, 2006; Kirk-Kuwaye & Kirk-Kuwaye, 2007; Marti, 2007; Peter &

Cataldi, 2005; Sturtz, 2006; Winter, Harris & Ziegler, 2001). The majority of these researchers have selected samples based upon high school graduation year rather than on first semester of college attendance. The benefit of this definition of a population is that every student had the potential to attend an institution of higher education at the same time. An important aspect in the observation of multiple pathways through higher education is that not all students start their journey at the same time, so defining a population by the point in time in which they could have started provides a common point of departure for analysis.

Some of the researchers also limited their samples to those students who have attended a four-year institution at any time during their observation period. Adelman (2006a) argued that to study baccalaureate degree attainment, one should consider only those who attended a baccalaureate granting institution. He argued that the best indicator of intent to earn a baccalaureate is using academic records of those who ever attended a baccalaureate degree granting institution and not self-reported survey data. Those authors not using this restriction (Marti, 2007; Bahr, 2009; Winter, Harris & Ziegler, 2001) typically were only studying community college pathways.

Reason (2009) concluded that the study of student persistence should be an institution specific enterprise because persistence is affected by “organizational context and the local student peer environment” (p. 678). Metz (2004) and Tinto (1993) also support the importance of institutional based studies over national studies in understanding of student retention and persistence. To apply this concept to multi-institutional studies then would require looking at a cluster of institutions where students migrate easily.

Period of observation. Another legacy of the Student Right-to-Know (1996) cohort definitions is that the federal government requires four-year and six-year graduation reporting. These periods have also become standards in higher education research. As with sample selection, the works of Bahr (2009) and Li (2010) serve as examples of these standards. Research using one of the national datasets of postsecondary students, such as Adelman (1999 – 2006) and Goldrick-Rab (2006), tend to have longer periods of observation of 12 to 13 years. These longer periods of observation yield more complete data about student pathways and graduation. However, researchers still struggle to accommodate the portions of their samples still enrolled at institutions of higher education at the end of their observation period. Any artificial period of observation will always lead to some level of censoring of the data set. Another challenge is determining if students who do not complete a degree during the observation period are dropouts (students who never return) or stop-outs (students who will eventually return). These differentiations cannot be made based on transcript data alone when the observation period is too short.

Variables included. The variables included in any analysis also affect the study of multi-institutional attendance. Adelman (2006a) made the argument that the best data is not self-reported data, which is open to interpretation, but transcript data, which tracks not what a student intends or wants to do, but what actually happened. While the researcher cannot always interpret student choice using transcript data, actual pathways are mapped without uncertainty. Variables used in pathway determination include: number of institutions attended (Bach et al., 2000; Goldrick-Rab, 2006; Peter & Cataldi, 2005), continuity of enrollment (Goldrick-Rab, 2006), and type of institution attended (Bach et al., 2000; Peter & Cataldi, 2005).

The interaction of variables and statistical tests used is also important. Certain statistical tests require converting continuous data to nominal data thus potentially losing some of the information. In each of the studies using national data sets (Appendix A), researchers had to nominalize some portion of their data in order to use their chosen statistical method. In each of Adelman's studies, the number and type of institutions attended were included in analysis, but the timing and nature of credit accumulation was lost in order to create variables that could be analyzed in a logistic regression.

Statistics. Appendix A provides a compendium of studies of enrollment pathways. All of the studies in Appendix A were conducted, in part, by using a large descriptive analysis of multi-institutional pathways. The analyses were more than just a brief presentation of variables however. Each study used national datasets and aggregated students in groups based upon their own pathway definitions. Adelman (1999 – 2006) frequently used the order of institutional attendance as the basis for his attendance pathways. Once the definition of attendance pathways was complete, however, all of the authors used simple descriptive statistics and occasionally used t-tests to find differences within those groups.

All but two of the studies (Alfonso, 2006; Winter, Harris & Zeigler, 2001) using inferential statistical techniques relied on regressions. Logistic regression is a popular choice among educational researchers because it is one of the few statistical tests that allows for a dichotomous outcome variable (Cabrera, 1994; Peng, So, Stage & John, 2002). Many of the outcomes in higher education enrollment studies are dichotomous:

deciding to attend college, deciding to come back each semester, deciding to transfer, graduation. As Cabrera points out "there are no interval scales to describe such behaviors. Either an individual attends college or not, majors in a hard sciences or not, stays or leaves the institution, or obtains a bachelor degree or not" (p. 225).

Fundamentally, logistic regression "seeks to identify a combination of [independent variables]…that best predicts membership in a particular group, as measured by a categorical [dependent variable]" (Mertler & Vannatta, 2005, p. 313). In addition to the ability to use a dichotomous (categorical) variable as the dependent variable, the use of logistic regression also has the following strengths:

1. Independent variables can be continuous, discrete or dichotomous, or a mix thereof.
2. Predictor variables do not have to be normally distributed, linearly related, or have equal variances within each group.
3. Statistic can have more than two outcomes, which may or may not have order; and
4. Statistic can produce non-linear models.

(Mertler & Vannatta, 2005; Peng, So, Stage & John, 2002; Tabachnick & Fidell, 2007). Lastly, regression techniques predict the *outcome* of a student on a particular pathway, in most cases whether or not a given student will graduate. However, it is more useful to identify *when* in the pathways students are most at risk. These points in time can be the areas of focus for those interested in improving retention.

A general challenge for any statistic employed by higher education researchers is the occasional necessity to create dummy variables. For example, McCormack (2003) reduced the age of students to two groups: those twenty-four years of age and older and those less than twenty-four years of age. This creates artificial groups, the hazard being what evidence exists to indicate that the break should be between the age of twenty-three and twenty-four? Or just two groups? Another example is a study by Adelman (1999) where the number of schools attended was coded as one school versus more than one. These researchers failed to cite the theoretical or statistical reasoning for choosing these break points.

New Frontiers for Analysis of Multi-Institutional Attendance

Reason (2009) argues that as a direct result of the increasing complexity of the models of student persistence and student attendance pathways "higher education researchers will need to become more facile with...advanced analytic and design techniques" (p. 676). In the last two decades educational researchers have begun to experiment with more sophisticated analytical techniques, techniques in some cases that have been used successfully in other fields for a while. One in particular shows promise in helping to better understand the effects of multi-institutional attendance pathways: survival analysis.

Survival analysis evaluates the time it takes for something to happen, in the case of multi-institutional attendees leaving college, whether because of graduation or because of dropping out. There are two types of survival analysis. The first is where survival analysis is used to "describe the proportion of cases surviving at various times, within a single group or separately for different groups" (Tabachnick & Fidell, 2007, p. 507). The second type is used to "assess the relationship between survival time and a set of covariates (predictors), with treatment considered one of the covariates, to determine whether treatment differences are present after statistically controlling for the other covariates" (Tabachnick & Fidell, 2007, p. 507). Where survival analysis has been employed in higher education, it has been typically with the former model (Bahr, 2008; Calcagno, Crosta, Bailey, & Jenkins, 2007; Willett & Singer, 1991). Bahr (2008) best illustrates the difference between logistic regression and survival analysis by stating, "[survival analysis] provides information about the likelihood of the occurrence of the event per unit of time, while [logistic regression] provides information about the likelihood of occurrence over the entire observation period" (p. 12). Calcagno, Crosta, Bailey and Jenkins (2007) articulate that while there are many studies using longitudinal data, most researchers are really only looking at two points in time: when students start their postsecondary education and after a period in time where most students have been given a chance to graduate. They argue that while researchers try to estimate the direct effect of their observations, "their strategy masks fundamental variation that explains degree completion because factors such as enrollment patterns or institutional characteristics are likely to change over time" (p. 779). Willett and Singer (1991) argue that "rather than asking whether students dropout before the end of their senior year, they should ask: When are they at the greatest risk of dropping out?" (p. 408).

There are several reasons that make survival analysis an appealing tool for researchers studying multi-institutional attendance. First is that the survival time (dependent variable) can be unknown for many of the cases (Tabachnick & Fidell, 2007). These cases have what is referred to as *censored* event times (Tabachnick & Fidell, 2007; Willet & Singer, 1991, 1993, & 1995). In their series of articles on survival analysis, Willett and Singer (1991, 1993, & 1995) show that to answer "When?" types of research questions using conventional statistical techniques, researchers must know the outcome for every person in the study. In most studies of institutional attendance, there are always students who are still enrolled, a proportion that have not yet graduated but are still attending an institution of higher education. These cases have long confounded researchers, and most researchers have had to discard these cases, usually by only sampling students who have graduated.

Secondly, using survival analysis prevents artificial dichotomization of the data set, as is required to use logistic regression (Willett & Singer, 1991). Willett and Singer argue that “sample splitting eliminates potentially meaningful variation in event times by clustering together everyone” (p. 408) into artificial groups. They further argue that dichotomy-based analyses of student pathways “are likely to become increasingly invalid in the 1990s and beyond as once nontraditional education trajectories become the norm” (p. 409). They conclude that dichotomizing outcome data does not resolve the censoring dilemma but simply obscures it from view.

There are several examples of survival analysis being applied to educational related questions. Willett and Singer (1991) examined when teachers are most likely to leave the profession and when students are most at risk for leaving college. They further expanded this question in 1995 using multiple-spell discrete-time survival analysis to examine “students’ and teachers’ entries into, and exits from, school” (p. 41). Ronco (1995) examined the risk associated with graduating, transferring, or dropping out of school of a sample of 1,635 first-time, full-time freshmen. Murtaugh, Burns, and Schuster (1999) examined the risk of a student’s probability of leaving school of a cohort of 8,867 first-time, full-time freshmen using demographic and academic predictors. Calcagno, Crosta, Bailey and Jenkins (2007) used a cohort of 42,641 first-time community college students to examine the between-group differences of older and younger students. Chimka, Reed-Rhoads, and Barker (2007-2008) examined the risk of graduation among a cohort of 429 first-time, full-time engineering students. Lastly, Visser, Luwel, and Moed (2007) examined the propensity to attain a Ph.D. at any of the five largest Flemish universities.

Survival analysis provides a tantalizing option to study the phenomenon of multi-institutional attendance. Its strengths include the ability to use censored data as well the ability to collect multiple data points during the period of observation, both of which are important characteristics of the phenomena in question. However, before it should be recommended as a statistic of choice, its use should be compared to the current tools used to study these phenomena in order to determine if the use of a more advanced and complicated statistic is warranted.

Conclusions

Many researchers have taken steps to describe the multi-institutional pathways of students and the observations are wide and varying but draw strength from the use of national datasets. The challenge that remains is two-fold. First is how to analyze the data generated by these complicated pathways. These data are as multi-dimensional as the student pathways, and do not lend themselves well to traditional statistical analyses. Second, the observations made from these national studies have not yet informed new methods for the study of higher education. While many of these studies establish new standards of sample selection, a few still hold tight to the framework of the study of first-time, full-time cohorts. There is often a disconnection between the phenomena being studied and the methods being used to study it. The magnitude of that disconnect is, as of yet, uncertain. Novel statistical methodologies present a method of studying these phenomena which may prove to yield more information about the phenomena in question, that of multi-institutional pathways and their effect on student success.

Chapter Three: Research Methodology

Introduction

The purpose of this research was to compare different statistical approaches to studying the phenomenon of multi-institutional attendance. The increasing complexity of student attendance patterns in higher education has stretched the limit of common research methods (Reason, 2009). The challenge for practitioners is to make data driven decisions in the most efficient and effective manner possible. In order to accomplish this, a better understanding of the most appropriate statistical method(s) is required, specifically the ability of each method to provide the desired data and the practicality of applying each method. The goal of this research was to compare the application of three statistical methods: logistic regression, discriminant analysis, and survival analysis to the study of multi-institutional attendance.

Research Design and Analysis

Research question. How does each statistical methodology (logistic regression, discriminant analysis, and survival analysis) affect the sample selection, periods of observation, assumptions, variables used and outcomes (what questions can be answered) in a study of multi-institutional attendance? Secondarily, how do administrators ensure that they are evaluating and correctly applying persistence and retention research to their respective institutions?

Each individual statistic also has a unique research question.

- Logistic regression: "Can the obtainment of a baccalaureate degree be correctly predicted from a student's attendance pattern?"

- Discriminant Analysis: “Are there significant differences between those who obtain a baccalaureate and those that do not?”
- Survival Analysis: “How does the nature and timing of student enrollment patterns affect the timing of baccalaureate degree attainment?”

Participants. The participant pool for this study includes all students who graduated from high school the spring of 2000 who later attended at least one semester as a baccalaureate degree-seeking student at the University of Nevada, Reno. The participant pool includes new freshmen and transfer students, as well as completers and non-completers.

Sample. A random sample of 600 was selected from the population pool using a table of random numbers. After testing the data for outliers, 16 cases were discarded leaving a remaining sample of 584. This sample was large enough to provide a 0.95 confidence interval for each statistic (Mertler & Vannatta, 2005; Tabachnick & Fidell, 2007).

Data collection. Data for each participant was extracted from the institution’s student information system, collected from submitted transcripts, and participants were matched to the National Student Clearinghouse to obtain a complete record of institutional attendance. The observation period included all postsecondary enrollments for the sample between 2000 and either 2010 or the time of their first baccalaureate, whichever was sooner. Once all data was collected, all participants were de-identified.

Dependent variable. The dependent variable in each analysis was graduation status at the end of the observation period. This variable had three possible values: graduation (first baccalaureate), still enrolled, and neither enrolled nor graduated. Any enrollment that occurs after the obtainment of the first baccalaureate degree was excluded from this study.

Independent variables. Different variables were collected and/or calculated to best fit the application of each statistic. The independent variables were derived from the following data to best describe course-taking patterns at each institution attended and to predict success. Course taking patterns are generically represented by indicators such as number of units both attempted and earned each semester at each institution attended. The bio/demographic variables (gender and ethnicity) are included in order to look for group differences in outcomes. The following variables below are commonly used in research as reasonable considerations to predict student success (Adelman 1999, 2006).

- High School GPA
- In-state or Out-of-State
- Enrollment Status (Less than Half Time, Half-Time, Three-Quarter Time or Full Time) per Semester (fall, spring and summer) per Institution Attended
- Number of Institutions Attended per Semester
- Gender
- Ethnicity

Time was also an important consideration in this study. One of the goals of the study was to compare the results from several statistics, and at least one, survival analysis, takes into account specific periods of time and the outcome at the end of each of those periods. For this study, time was measured by academic terms with each year consisting of three terms; fall, spring and summer. As different institutions have different periods of enrollment (semester, trimesters, etc.), the enrollment periods obtained through the

Clearinghouse were normalized to a standard semester system, where August through December was coded as fall, January through May was coded as spring, and May through August was coded as summer. Because summer is normally not considered a regular semester, two time sequences were used. The first sequence considered enrollment in all available semesters (in this case 30 over a ten-year period) versus enrollment in just those semesters considered regular (in this case 20).

In order to take into account term based observations in statistics designed to only have one observation per predictor variable per case, it was necessary to create a series of derived variables. Each of the following variables was developed to represent enrollment patterns and was used only in the multinomial regression or discriminant analysis (Table 3.1). Survival analysis allowed time variables to be taken into consideration in a different manner. Each value was chosen to represent intensity of enrollment. A student enrolled every semester would have a *ratio of total enrolled semesters* equal to 1.0; a student enrolled in only one or two semesters would have a value close to 0. The six variables consisted of three that included summer *ratio of total enrolled semesters, ratio of total full-time enrolled semesters, and ratio of total semesters as a multi-institutional attendee (MIA)* and three that did not include summer *ratio of enrolled regular semesters, ratio of full-time enrolled regular semesters and ratio of regular semesters as a multi-institutional attendee (MIA).*

Table 3.1

List of variables used for each statistic.

	Logistic Regression	Discriminant Analysis	Survival Analysis
Dependent Variable	• Status at conclusion of observation period • Bachelor degree • Still Enrolled • No degree/no enrollment	• Status at conclusion of observation period • Bachelor degree • Still Enrolled • No degree/no enrollment	• Status at conclusion of observation period • Bachelor degree • Still enrolled • No degree/no enrollment
Independent Variables	• High School GPA • Overall Number of Schools Attended • Number of semesters with multi-institutional enrollment • Ratio of Total Enrolled Semesters OR Ratio of Enrolled Regular Semesters • Ratio of Total Enrolled Full Time Semesters OR Ratio of Enrolled Regular Full Time Semesters • Ratio of Total Semesters as a Multi-institutional attendee OR Ratio of Regular Semesters as a Multi-institutional Attendee • Gender • Ethnicity	• High School GPA • Overall Number of Schools Attended • Number of semesters with multi-institutional enrollment • Ratio of Total Enrolled Semesters • Ratio of Total Enrolled Full Time Semesters • Ratio of Total Semesters as a Multi-institutional Attendee • Gender • Ethnicity	• High School GPA • Number of semesters enrolled • Overall Number of Schools Attended • Number of semesters with multi-institutional enrollment • Summer enrollment • Gender • Ethnicity

Data Analysis

Descriptive data. The variables used in this study were reported using descriptive statistics. Enrollment patterns, aggregated and on a semester by semester basis, were summarized using frequencies, measures of central tendency, distributions and cross tabulations. Data were reported for the entire sample and by each group to look for any differences between the two groups.

Logistic regression. Logistic regression was chosen for this study because it is the most popular method of analyzing student outcomes. All but two of the studies in Appendix A (Alfonso, 2006; Winter, Harris & Zeigler, 2001) using inferential statistical techniques relied on regression tests. Logistic regression has been a popular choice among educational researchers because it is one of the few statistical tests that allows for a dichotomous outcome variable (Cabrera, 1994; Peng, So, Stage & John, 2002).

As outlined in Mertler and Vannatta (2005), an initial regression was conducted using SPSS to calculate the Mahalanobis distance in order to identify any potential outliers and to examine multicollinearity among the predictor variables identified in Table 3.1. Outliers identified were removed from analysis. A multinomial logistic regression was then performed and the resulting model was assessed using the following goodness of fit indices: -2 Log Likelihood, Goodness of Fit, and Model Chi-Square with degree of freedom and level of significance. The percentage of correctly classified cases was reported.

Discriminant analysis. Discriminant analysis was the second statistic chosen for this study because it is the parametric equivalent of logistic regression (Mertler & Vannatta, 2005), thus still allowing for a categorical outcome variable. Its primary difference is that discriminant analysis is a parametric statistic, which requires the user to meet more assumptions about the nature of the data than a logistic regression requires.

Univariate normality, homoscedasticity, and linearity were assessed with the use of bivariate scatterplots. Transformations, including a natural log and logarithmic transformation, were attempted in order to reduce the skewness of several variables. Additionally, discriminant analysis can only utilize non-categorical variables, so ethnicity and gender, both used in logistic regression, were removed from this analysis. The discriminant analysis was conducted and resulting discriminant functions were evaluated for significance, and then chosen for inclusion and interpreted. The accuracy of the classification was reported.

Survival analysis. Survival analysis was the third statistic chosen for this study. Survival analysis had been proposed by a few researchers as a potential tool to better study complex enrollment patterns such as those encountered when studying multi-institutional attendance (Ronco, 1995; Willett & Singer, 1991).

An initial regression was conducted using SPSS on the prediction group to calculate the Mahalanobis distance in order to identify any potential outliers. Multivariate normality, homoscedasticity, and linearity were assessed with the use of bivariate scatterplots. While Tabachnik and Fidell (2007) note that survival analysis is robust against violations of these assumptions, they also argue that meeting them increases the power and prediction of a survival analysis. A Cox regression survival analysis as described by Tabachnik and Fidell (2007) was conducted utilizing SPSS to obtain the overall risk score for survival time and the overall survival function.

Comparative analysis. The last step in this study was to comparatively evaluate all three methods on a series of criteria. Each statistic was evaluated on the following points: treatment of the raw data, treatment of the cases (did some have to be dismissed from the study and why?), which variables were included and which were not, the appropriateness of the question asked in the study of multi-institutional attendance, and the limitations of each statistic. Each of these points was charted and evaluated. The framework for this evaluation used three criteria. The first was usefulness of the individual statistic in answering questions related to multi-institutional attendance. The second was ease of use of the statistic and the clarity of the interpretation of its results. Lastly was the ability of the statistic to use as many of the cases and variables without significant transformations or exclusions as possible.

Summary

Three different statistical tests, including logistic regression, discriminant analysis, and survival analysis, were conducted on the same data set. Additionally a descriptive analysis was conducted in order to describe general trends within the data. The objective of these three analyses was not necessarily to identify factors influencing baccalaureate degree attainment, but to look at the outcomes of each statistic and its appropriateness for use in the study of multi-institutional attendance. A comparative analysis of the process used for each statistic was completed and the use of each statistic assessed for its appropriateness to be applied to multi-institutional attendance patterns. In chapter four, descriptive statistics and each statistical analysis are presented and then a comparison of the three is presented in chapter five.

Chapter Four: Results

Introduction

The goal of this study was to investigate the similarities and differences that result from application of three different multivariate statistics to the phenomenon of multi-institutional attendance. Using SPSS (Version 20), a multinomial logistic regression, a discriminant analysis, and a survival analysis were performed on data from a sample of students selected from the high school 2000 graduation class who attended the University of Nevada, Reno. Each student attended the university for at least one semester as a degree-seeking student during the ten-year period from 2000 through 2010. That is, each student in the sample attended University of Nevada, Reno at least one semester during the identified ten-year period.

This chapter is divided into five sections. Section one contains descriptive statistics. In section two the results of the multinomial logistic regression are reported, in section three the results of the discriminant analysis are reported, and in section four the results of the survival analysis are reported. Section five provides a summary of the research findings.

Descriptive Statistics

The base data file used for all analyses contained one record per case per enrollment period per institution. For a student who attended two institutions during the same semester, two records would exist, one for each institution. Each record indicated the level of the student's enrollment (less than half-time, half-time, three-quarter time and full-time). Enrollment periods are defined by the reporting institutions when they submit term files to the National Student Clearinghouse (Clearinghouse). Each student enrollment level was converted to the following numeric values for analysis purposes; the conversion is illustrated in Table 4.1.

Table 4.1

Load Conversions

Clearinghouse Enrollment Load	Enrollment Load as Applied in Study
Less than Half Time	0.25
Half Time	0.50
Three-Quarter Time	0.75
Full Time	1.00

This conversion allowed *enrollment load* to be treated as a continuous variable. For example, values could be summed to indicate one aggregate value when a student attended more than one institution in the same enrollment period. If a student attended two institutions during the same enrollment period, one reported as less than half time (0.25) and the other reported as half time (0.50), then the students combined load for the enrollment period would be three-quarter time (0.75). Lastly, the term of graduation was also obtained from the Clearinghouse. Any enrollment after the first baccalaureate was earned from any institution was removed from the file prior to analysis. The data from the Clearinghouse were combined with demographic data obtained through the institutional student information system: gender, ethnicity, high school state, and high school GPA.

Out of the original 600 students in the sample, 16 cases were discarded prior to analysis because they were identified as outliers, leaving the total sample size as N = 584. These cases were identified by calculating the Mahalanobis distance scores for each case in SPSS (Version 20). For this study the critical value was found to be χ^2 (8) = 26.13.

There were 16 cases with a Mahalanobis distance score greater than the critical value and thus were removed from analysis.

Out of the 584 cases, 447 (76.5%) had graduated by the end of the 10-year study period, 27 (4.6%) were still enrolled, and 110 (18.8%) had neither graduated nor were enrolled at the end of ten years. Table 4.2 provides a summary of frequencies, percentages, and cumulative percent for the outcome variable, *graduation*.

Table 4.2

Frequency of Outcome Variable

	Frequency N	Percent %	Cumulative Percent %
Graduated	447	76.5	76.5
Still Enrolled	27	4.6	81.2
Not Enrolled/Not Graduated	110	18.8	100.0
Total	584	100.0	

The distribution for the *gender* variable was 344 (58.9%) female and 240 (41.1%) male (Table 4.3). When examining the gender distribution by outcome, the distribution for each level holds true except for those who were observed to be not enrolled and not graduated by the end of the observation period. Of that group, 62 (56.4%) were female and 48 (43.6%) were male.

Table 4.3

Gender Distribution by Outcome

		Gender				Total
		Female		Male		
		N	%	N	%	
Outcome	Graduated	266	59.5	181	40.5	447
	Still Enrolled	16	59.3	11	40.7	27
	Not Enrolled/Not Graduated	62	56.4	48	43.6	110
Total		344	58.9	240	41.1	584

The distribution for the *ethnicity* variable for those that graduated was similar to the overall sample (Table 4.4). There were, however, differences in the distribution of ethnicity in those still enrolled or neither enrolled nor graduated. There were also a large number of cells with zero frequency, which is a dilemma for multivariate analyses.

Table 4.4

Ethnicity Distribution by Outcome

		Ethnicity													
		Non-Resident Alien		American Indian/ Alaskan Native		Asian/ Pacific Islander		Black		Hispanic/ Latino		Unknown		White	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%
Outcome	Graduated	6	1.3	5	1.1	25	5.6	8	1.8	21	4.7	16	3.6	366	81.9
	Still Enrolled	0	0	0	0	1	3.7	0	0	2	7.4	0	0	24	88.9
	Not Enrolled/ Graduated	1	0.9	1	0.9	8	7.3	3	2.7	5	4.5	7	6.4	85	77.3
Total		7	1.2	6	1.0	34	5.8	11	1.9	28	4.8	23	3.9	475	81.3

Because of the large number of cells with a zero value (Table 4.4), ethnicity was recoded into white or non-white for each analysis (Table 4.5). The new *ethnicity* variable was 475 (81.3%) white and 109 (18.7%) non-white. While the *ethnicity* distribution of those who had graduated was similar to the overall sample, white students were more likely to be *still enrolled* and less likely to be *neither enrolled nor graduated* than non-white students.

Table 4.5

Recoded Ethnicity Distributions by Outcome

		Ethnicity				Total
		White		Non-White		
		N	%	N	%	
Outcome	Graduated	366	81.9	81	18.1	447
	Still Enrolled	24	88.9	3	11.1	27
	Not Enrolled/Not Graduated	85	77.3	25	22.7	110
Total		475	81.3	109	18.7	584

While the original sample included the specific state in which each student attended high school, the number of cells with a zero value was high. *High school state* was recoded into a variable with just two values, in-state, for Nevada high school graduates, and out-of-state, for non-Nevada high school graduates. For this sample, 500 (85.6%) were from an in-state high school while 84 (14.4%) were from out-of-state (Table 4.6).

Table 4.6

High School State by Outcome

		High School State				Total
		Out-of-State		In-State		
		N	%	N	%	
Outcome	Graduated	66	14.8	381	85.2	447
	Still Enrolled	3	11.1	24	88.9	27
	Not Enrolled/Not Graduated	15	13.6	95	86.4	110
Total		84	14.4	500	85.6	584

For the sample, 514 (88%) attended the University of Nevada, Reno as their first post-secondary institution, while 70 (12%) transferred to the University of Nevada, Reno from another institution (Table 4.7). However, of the 447 students in the sample who graduated, all attended the University of Nevada, Reno first. This may be due to sampling errors as the University of Nevada, Reno does graduate students who transfer in from other institutions and have no prior enrollment from Nevada.

Table 4.7

UNR First School by Outcome

		Is UNR the First School				Total
		No		Yes		
		N	%	N	%	
	Graduated	0	0	447	100%	447
Outcome	Still Enrolled	13	48.1%	14	51.9%	27
	Not Enrolled/Not Graduated	57	51.8%	53	48.2%	110
Total		70	12.0%	514	88.0%	584

The distribution of *high school GPA* ranged from 1.64 to 4.00, with a mean of 3.45 (Figure 4.1). The distribution of this predictor variable was affected by the minimum GPA for regular freshmen admission at the university being 3.00. This gave the *high school GPA* its overall moderate skew of -0.687. Figure 4.1 provides a plot of the frequency distribution of GPA values, with a curve of best-fit illustrating the moderate negative skew of the distribution of this variable.

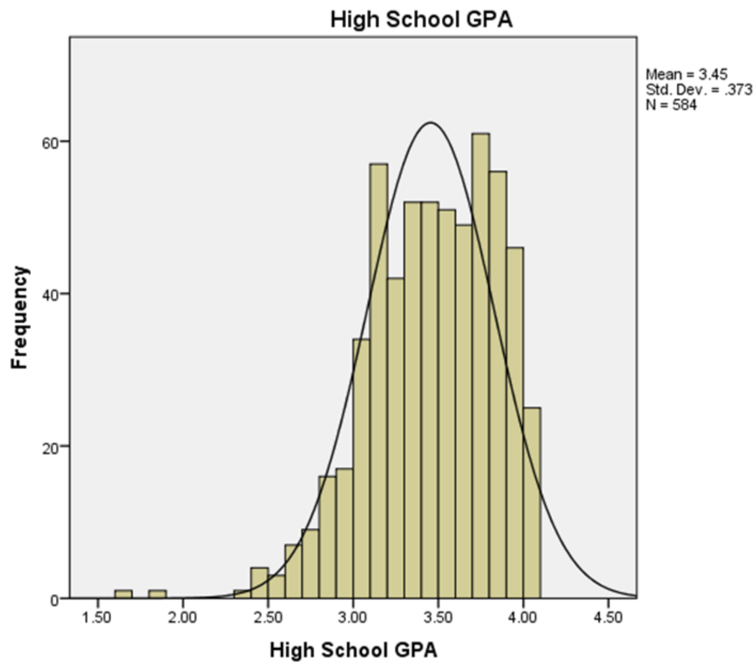

Figure 4.1 *Distribution of High School GPA*

The distribution for the *number of post-secondary institutions* ranged from one to six institutions with 240 (41.1%) attending only one institution and 344 (58.9%) attending between two and six institutions either prior to graduation or prior to the end of the observation period (Table 4.8). The mean number of institutions attended by the students in the sample is 1.8 (Figure 4.2). Another way to state this distribution is that 240 (41.1%) attended only the University of Nevada, Reno during the observation period and 344 (58.9%) attended the University of Nevada, Reno and at least one other institution prior to graduation or the end of the observation period.

Table 4.8

Number of Post-secondary Institutions Attended by Outcome

		Total Number of Post-secondary Institutions												
		1		2		3		4		5		6		Total
		N	%	N	%	N	%	N	%	N	%	N	%	
Outcome	Graduated	198	44.3	183	40.9	52	11.6	11	2.5	2	0.4	1	0.2	447
	Still Enrolled	3	11.1	15	55.6	5	18.5	4	14.8	0	0.0	0	0.0	27
	Not Enrolled/Not Graduated	39	35.5	48	43.6	19	17.3	4	3.6	0	0.0	0	0.0	110
Total		240	41.1	246	42.1	76	13.0	19	3.3	2	0.3	1	0.2	584

Figure 4.2 provides a plot of the frequency distribution of *number of post-secondary institutions* attended, overlain by a curve of best fit. The mean *number of post-secondary institutions* is 1.8, with a standard deviation of 0.831. This predictor variable ranged from one to six.

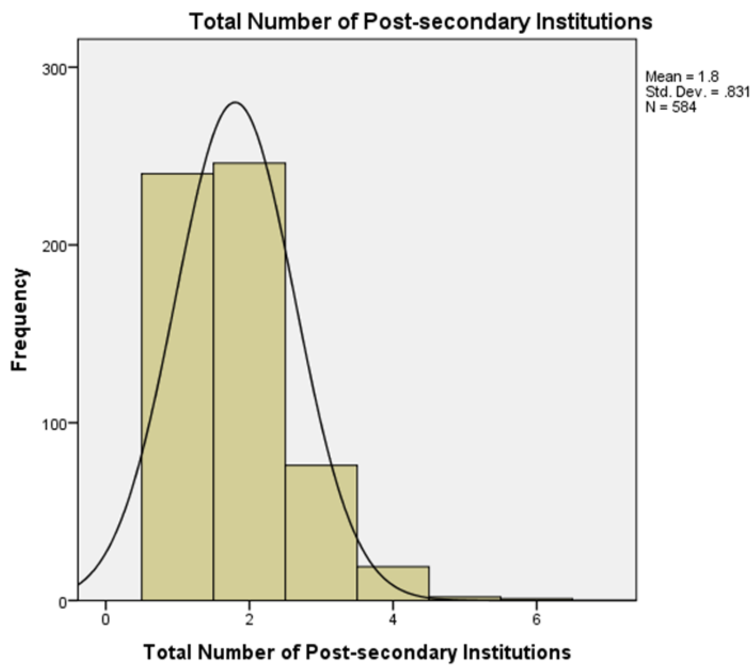

Figure 4.2 *Distribution of Number of Post-secondary Institutions*

As a component of the data analysis, six derived variables were developed to represent enrollment patterns and used in the multinomial regression and discriminant analysis. The six derived variables include three that take into account summer terms: *ratio of total enrolled semesters, ratio of total full-time enrolled semesters, and ratio of total semesters as a multi-institutional attendee (MIA)* and three that do not take into account summer terms: *ratio of enrolled regular semesters, ratio of full-time enrolled regular semesters and ratio of regular semesters as a multi-institutional attendee (MIA).* Table 4.9 lists the minimum and maximum, the mean, standard deviation and skewness for each variable. In this study, each variable ranged from 0 to 1. All variables were moderately to significantly skewed.

Table 4.9

Descriptive Statistics of Six Enrollment Measures

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness
Ratio of Enrolled Regular Semesters	584	.050	1.00	.830	.252	-1.392
Ratio of Total Enrolled Semesters	584	.035	1.00	.635	.205	-1.150
Ratio of Full Time Enrolled Regular Semesters	584	.050	1.00	.745	.300	-.918
Ratio of Total Full Time Enrolled Semesters	584	.035	.909	.539	.221	-.856
Ratio of Regular Semesters MIA	584	.000	.444	.031	.070	2.751
Ratio of Total Semesters MIA	584	.000	.333	.023	.051	2.886

The last predictor variable, *number of total semesters as a multi-institutional attendee*, measured the number of semesters a student was enrolled in more than one institution. Out of the 584 students in the study, 450 (77.1%) were never enrolled in more than one institution at the same time (Table 4.10). Conversely, 134 (22.9%) of students in the study were a multi-institutional attendee at least once. Table 4.10 provides a summary of multi-institutional attendance by outcome. For each outcome, the total number and overall percentage of students is presented by the total number of semesters they attended more than one institution.

Table 4.10

Multi-institutional Attendance by Outcome

		Number of Total Semesters MIA												Total
		0		1		2		3		4		5		
		N	%	N	%	N	%	N	%	N	%	N	%	
Outcome	Graduated	340	76.1	66	14.8	26	5.8	7	1.6	7	1.6	1	0.2	447
	Still Enrolled	21	77.8	4	14.8	2	7.4	0	0.0	0	0.0	0	0.0	27
	Not Enrolled/ Graduated	89	80.9	14	12.7	4	3.6	1	0.9	1	0.9	1	0.9	110
Total		450	77.1	84	14.4	32	5.5	8	1.4	8	1.4	2	0.3	584

Out of those students who did qualify as an MIA, the majority (62.7%) only did so for one semester (Table 4.11). An additional 23.9% attended more than one institution for two semesters. There were eight students (6.0%) who attended more than one institution for either three or four semesters. Lastly 1.5% or two students attended more than one institution for five semesters.

Table 4.11

Number of Total Semesters MIA

		Frequency N	Percent %	Valid Percent %	Cumulative Percent %
Number of Semesters MIA	1	84	14.4	62.7	62.7
	2	32	5.5	23.9	86.6
	3	8	1.4	6.0	92.5
	4	8	1.4	6.0	98.5
	5	2	.3	1.5	100.0
	Total	134	22.9	100.0	
No Semesters MIA		450	77.1		
Total		584	100.0		

Logistic Regression

The first statistic explored was a multinomial logistic regression. Multinomial logistic regression is used for predicting the outcome of a categorical dependent variable based on multiple predictor variables. Multinomial logistic regression is a commonly used statistic for the study graduation in higher education.

Two separate multinomial logistic regression models were tested; the first model included enrollment from summer terms, while the second model did not. Both models included the following predictor variables: *gender, ethnicity, in-state high school, and UNR first school*. In addition to these variables, the first model contained the measurements of enrollment that included summer terms r*atio of total enrolled semesters, ratio of total full-time enrolled semesters,* and *ratio of total semesters as a multi-institutional attendee (MIA).* The second model included three predictor variables that did not include summer *ratio of enrolled regular semesters, ratio of full-time enrolled regular semesters and ratio of regular semesters as a multi-institutional attendee (MIA).* Regression results indicate that both models were statistically reliable in distinguishing between students who graduated from those that were either still enrolled or not graduated nor enrolled.

Model 1. The first model, hereafter referred to as the Summer Model included all enrollment terms had -2 Log Likelihood = 135.400, Goodness-of-Fit = 135.400, and χ^2 (18) = 636.876, and p < .001 (Table 4.12).

Table 4.12

Model Fitting Information – Summer Model

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	772.276			
Final	135.400	636.876	18	< .001

The Summer Model correctly classified 94.9% of the cases (Table 4.13). The Summer Model had a high Nagelkerke Pseudo R-Square value indicating that a large amount of variation within the sample was explained by this model. The Summer Model's Nagelkerke R-Square value of 0.905 indicates this model explained nearly 91% of the variation.

Table 4.13

Classification Data – Summer Model

Observed	Predicted			
	Graduated N	Still Enrolled N	Not Enrolled/Not Graduated N	Percent Correct %
Graduated	445	0	2	99.6
Still Enrolled	2	7	18	25.9
Not Enrolled/Not Graduated	2	6	102	92.7
Overall Percentage	76.9%	2.2%	20.9%	94.9

The Likelihood Ratio test data for the Summer Model indicated that *high school GPA, number of schools attended, ratio of enrolled semesters, ratio of full-time enrolled semesters, ratio of semester with multi-institutional attendance, in-state high school* and *UNR first school attended* all have a significant contribution to the Summer Model (Table 4.14).

Table 4.14

Variable Contributions – Summer Model

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	135.400[a]	.000	0	.
HSGPA	156.958	21.559	2	< .001
SCHTOTAL	145.869	10.469	2	.005
RAT_ENRLS	166.540	31.140	2	< .001
RAT_FT_ENRLS	151.927	16.527	2	< .001
RAT_MULT_ENRLS	142.528	7.128	2	.028
GENDER	135.442	.043	2	.979
ETHNIC	198.034	62.634	2	< .001
HSINSTATE	150.286	14.886	2	< .001
FRSTSCHL	139.610	4.211	2	.122

Note. The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.
a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Wald statistics for the Summer Model indicated that *high school GPA, ratio of enrolled semesters, ratio of full-time enrolled semesters,* and *in-state high school* all significantly predicted graduation. Alternatively, *number of schools attended, ratio of enrolled semesters,* and *ratio of semesters of multi-institutional enrolled* were significant predictors of those who were still enrolled at the end of the observation period (Table 4.15). The beta values or coefficients indicated that as *high school GPA, ratio of enrolled semesters,* and *ratio of enrolled full-time semesters* increase, so does the likelihood of graduation. Additionally students from an in-state high school are also more likely to graduate.

Table 4.15

Parameter Estimates – Summer Model

Outcome[a]		B	Std. Error	Wald	df	Sig.
Graduated	Intercept	-37.056	9.411	15.504	1	< .001
	HSGPA	6.896	2.205	9.783	1	.002
	SCHTOTAL	-.699	.694	1.014	1	.314
	RAT_ENRLS	25.386	7.282	12.155	1	< .001
	RAT_FT_ENRLS	15.144	5.707	7.042	1	.008
	RAT_MULT_ENRLS	10.748	17.793	.365	1	.546
	[GENDER=0]	-.170	1.186	.020	1	.886
	[GENDER=1]	0[b]	.	.	0	.
	[HSINSTATE=0]	5.513	1.895	8.463	1	.004
	[HSINSTATE=1]	0[b]	.	.	0	.
	[FRSTSCHL=N]	-27.997	.000	.	1	.
	[FRSTSCHL=Y]	0[b]	.	.	0	.
	[Ethnic2=.00]	-1.462	1.512	.934	1	.334
	[Ethnic2=1.00]	0[b]	.	.	0	.
Still Enrolled	Intercept	-3.893	3.354	1.347	1	.246
	HSGPA	-.799	1.050	.579	1	.447
	SCHTOTAL	.936	.335	7.808	1	.005
	RAT_ENRLS	8.497	2.393	12.612	1	< .001
	RAT_FT_ENRLS	-2.864	2.820	1.031	1	.310
	RAT_MULT_ENRLS	-23.574	11.358	4.308	1	.038
	[GENDER=0]	-.091	.527	.029	1	.864
	[GENDER=1]	0[b]	.	.	0	.
	[HSINSTATE=0]	-.141	.794	.031	1	.859
	[HSINSTATE=1]	0[b]	.	.	0	.
	[FRSTSCHL=N]	.170	.816	.043	1	.835
	[FRSTSCHL=Y]	0[b]	.	.	0	.
	[Ethnic2=.00]	1.095	.746	2.158	1	.142
	[Ethnic2=1.00]	0[b]	.	.	0	.

Model 2. The second model, hereafter titled the Regular Model, included only regular enrollment terms, had -2 Log Likelihood = 141.242, Goodness-of-Fit = 141.242, and χ^2 (28) = 631.034, and p < .001 (Table 4.16).

Table 4.16

Model Fitting Information – Regular Model

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	772.276			
Final	141.242	631.034	18	< .001

The Regular Model correctly classified 94.9% of the cases (Table 4.17). The Regular Model also had a high Nagelkerke Pseudo R-Square value indicating that a large amount of variation within the sample was explained by this model. The Regular Model had a Nagelkerke R-Square value of 0.901, which indicates this model explained about 90% of the variation within the sample.

Table 4.17

Classification Data – Regular Model

Observed	Predicted			
	Graduated N	Still Enrolled N	Not Enrolled/Not Graduated N	Percent Correct %
Graduated	445	0	2	99.6
Still Enrolled	2	9	16	33.3
Not Enrolled/Not Graduated	4	6	100	90.9
Overall Percentage	77.2%	2.6%	20.2%	94.9

The Likelihood Ratio test data for the Regular Model indicated *high school GPA, number of schools attended, ratio of enrolled semesters, ratio of full-time enrolled* *semesters, ratio of semester with multi-institutional enrollment, in-state high school* and *UNR first school attended* all have a significant contribution to the regular model (Table 4.18).

Table 4.18

Variable Contributions – Regular Model

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	141.242a	.000	0	.
HSGPA	161.711	20.469	2	< .001
SCHTOTAL	151.016	9.773	2	.008
RAT_ENRL	168.062	26.819	2	< .001
RAT_FT_ENRL	170.559	29.317	2	< .001
RAT_MULT_ENRL	148.609	7.367	2	.025
GENDER	141.943	.701	2	.704
HSINSTATE	157.644	16.401	2	< .001
FRSTSCHL	193.010	51.767	2	< .001
Ethnic2	145.029	3.787	2	.151

Note. The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.
a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Similar to the Summer Model, Wald statistics for the Regular Model indicated that *high school GPA, ratio of enrolled semesters, ratio of full-time enrolled semesters* and *in-state high school* all significantly predict graduation. Alternatively, *number of schools attended, ratio of enrolled semesters,* and *ratio of semesters of multi-institutional enrolled* were significant predictors of students who were still enrolled at the end of the observation period (Table 4.19). The B-values or coefficients indicate that as *high school GPA, ratio of enrolled semesters,* and *ratio of enrolled full-time semesters* increase, so does the likelihood of graduation. Additionally, students from an in-state high school are also more likely to graduate.

Table 4.19

Parameter Estimates – Regular Model

Outcome[a]		B	Std. Error	Wald	df	Sig.
Graduated	Intercept	-34.666	8.593	16.274	1	< .001
	HSGPA	6.043	1.917	9.936	1	.002
	SCHTOTAL	-.103	.595	.030	1	.862
	RAT_ENRL	14.817	4.629	10.245	1	< .001
	RAT_FT_ENRL	13.583	4.207	10.422	1	< .001
	RAT_MULT_ENRL	4.310	10.986	.154	1	.695
	[GENDER=0]	.788	1.092	.520	1	.471
	[GENDER=1]	0[b]	.	.	0	.
	[HSINSTATE=0]	5.398	1.798	9.015	1	.003
	[HSINSTATE=1]	0[b]	.	.	0	.
	[FRSTSCHL=N]	-25.881	.000	.	1	.
	[FRSTSCHL=Y]	0[b]	.	.	0	.
	[Ethnic2=.00]	-1.208	1.313	.846	1	.358
	[Ethnic2=1.00]	0[b]	.	.	0	.
Still Enrolled	Intercept	-3.995	3.380	1.397	1	.237
	HSGPA	-.834	1.050	.631	1	.427
	SCHTOTAL	1.008	.340	8.809	1	.003
	RAT_ENRL	6.371	1.750	13.254	1	< .001
	RAT_FT_ENRL	-1.911	2.022	.894	1	.344
	RAT_MULT_ENRL	-18.225	8.257	4.872	1	.027
	[GENDER=0]	-.138	.529	.068	1	.795
	[GENDER=1]	0[b]	.	.	0	.
	[HSINSTATE=0]	-.244	.808	.091	1	.763
	[HSINSTATE=1]	0[b]	.	.	0	.
	[FRSTSCHL=N]	.285	.816	.122	1	.727
	[FRSTSCHL=Y]	0[b]	.	.	0	.
	[Ethnic2=.00]	1.059	.741	2.042	1	.153
	[Ethnic2=1.00]	0[b]	.	.	0	.

Summary. Overall, both models correctly classified over 90% of the cases. The Summer Model was able to classify students neither enrolled nor graduated more accurately while the Regular Model was able to classify students still enrolled more accurately. Additionally both models found the same variables to be significant: *high school GPA, number of schools attended, ratio of enrolled semesters, ratio of full-time enrolled semesters, ratio of semester with multi-institutional attendance, in-state high school* and *UNR first school attended*.

Discriminant Analysis

A descriptive stepwise discriminant analysis was planned for this study and conducted on the same data set as used in the logistic regression. However, due to non-normality of the *ratio of terms with multiple enrollment*, a natural log transformation was conducted. The results still contained a significantly non-normal distribution of this variable (Figure 4.3). According to Tabachnick and Fidell (1996), discriminant analysis is robust to violations of multivariate normality, as long as the violation is caused by skewness rather than outliers.

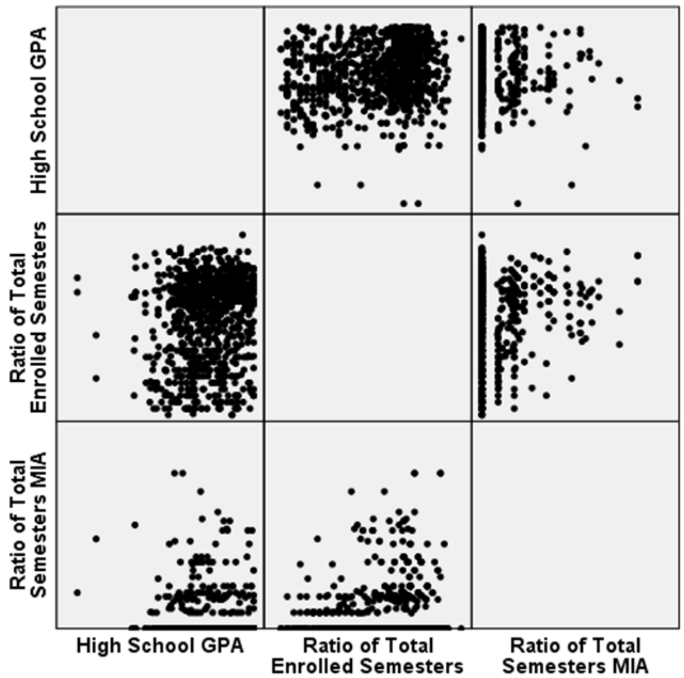

Figure 4.3 *Multivariate Scatter Plot*

Unlike the logistic regression, only one model was developed using the following variables: *in-state high school, high school GPA, total number of post-secondary institutions, ratio of total enrolled semesters, ratio of total full-time enrolled semesters,* and *ratio of total semesters MIA.* All predictor variables except *in-state high school* contribute to the identification of significant group differences (Table 4.20).

Table 4.20

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
In-State HS	.999	.168	2	581	.845
High School GPA	.948	15.948	2	581	< .001
Total Number of Post-secondary Institutions	.973	8.187	2	581	< .001
Ratio of Total Enrolled Semesters	.255	849.652	2	581	< .001
Ratio of Total Full Time Enrolled Semesters	.265	807.508	2	581	< .001
Ratio of Total Semesters MIA	.984	4.648	2	581	.010

The discriminant analysis generated two significant functions; Function 1 Λ = 0.215, χ^2 (4, N = 584) = 891.488, p < .001 and Function 2 Λ = 0.958, χ^2 (1, N = 584) = 24.710, p < .001 (Table 4.21) indicating that these predictor variables could accurately differentiate between students who graduated, were still enrolled or were no longer enrolled nor graduated. Even though both functions were significant, the second function contributes very little of the explanation of variance within the same. The high Wilks' value for function 2 indicates that nearly 96% of the variance is not explained by this function, and so it is not used for data interpretation.

Table 4.21

Discriminant Function Significance

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.215	891.488	4	< .001
2	.958	24.710	1	< .001

Overall, the discriminant function correctly predicted 91.6% of the cases included within the study (Table 4.22). Original classification results revealed that 96.6% of the *graduated* students were correctly classified, 18.5% of the *still enrolled* were correctly classified, and 89.1% of those *neither enrolled nor graduated* were correctly classified. Cross-validation accuracy dropped 0.02% of the classification accuracy.

Table 4.22

Discriminant Function Classification Results

		Outcome	Predicted Group Membership: Graduated	Predicted Group Membership: Still Enrolled	Predicted Group Membership: Not Enrolled/Not Graduated	Total
Original	Count	Graduated	432	11	4	447
		Still Enrolled	8	5	14	27
		Not Enrolled/Not Graduated	3	9	98	110
	%	Graduated	96.6	2.5	.9	100.0
		Still Enrolled	29.6	18.5	51.9	100.0
		Not Enrolled/Not Graduated	2.7	8.2	89.1	100.0
Cross-validated[b]	Count	Graduated	432	11	4	447
		Still Enrolled	8	5	14	27
		Not Enrolled/Not Graduated	3	10	97	110
	%	Graduated	96.6	2.5	.9	100.0
		Still Enrolled	29.6	18.5	51.9	100.0
		Not Enrolled/Not Graduated	2.7	9.1	88.2	100.0

a. 91.6% of original grouped cases correctly classified.
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
c. 91.4% of cross-validated grouped cases correctly classified.

Survival Analysis

With survival analysis, the assessment of all three outcomes is possible as it is with logistic regression and discriminant analysis (graduation, still enrolled, and neither enrolled nor graduated). The difference with survival analysis is that it takes into account the length of time (distinct periods of measurement) it takes to experience the outcome. The measurement of time used in this survival analysis was terms of enrollment. The survival analysis only included regular semesters of enrollment. By using all terms of enrollment, including summer, there were numerous gaps in the enrollment record for many of the students. This would be problematic within a survival analysis as survival analysis assumes that an observation, in this case enrollment, can be made for each time interval.

For each student record in the sample, the time variable, *semesters enrolled*, contained either the time the event of interest occurred (*graduation*) or, in censored cases, the last semester the student was enrolled. This variable is the total number of semesters the student was enrolled from 2000 through 2010. Each data record also contained values for the following predictor variables: *gender, ethnicity, in-state high school, high school GPA, number of colleges attended, number of semesters of multi-institutional enrollment, summer school enrollment,* and *UNR first school attended.* The first step was to complete a univariate analysis on the categorical predictor variables used within the analysis. This showed the shape of the survival function for each of the groups and indicated whether or not the groups have proportional hazards. This was useful for determining if there are significant differences between levels of each of the categorical predictor variables.

The only categorical predictor variables that were found to have significant univariate differences were number of colleges attended (Wilcoxon [Gehan] = 73.256, df = 4, $p < .001$) (Figure 4.4), and summer school enrollment (Wilcoxon [Gehan] = 3.917, df = 1, $p < 0.05$) (Figure 4.5). Pairwise comparisons of significance are presented in Table 4.23 for number of colleges attended. Because there were only two levels for summer school enrollment no pairwise comparisons exist. The median survival time for those attending summer school was 9.3 terms and 9.52 terms for those not attending summer school.

Figure 4.4 shows the survival function for the sample in groups by *number of colleges attended.* Each curve represents the probability of survival from one time period to the next. In general it can be stated that the probability of survival increases with the number of colleges attended. This is not necessarily indicative of a positive outcome as survival indicates that graduation has not occurred.

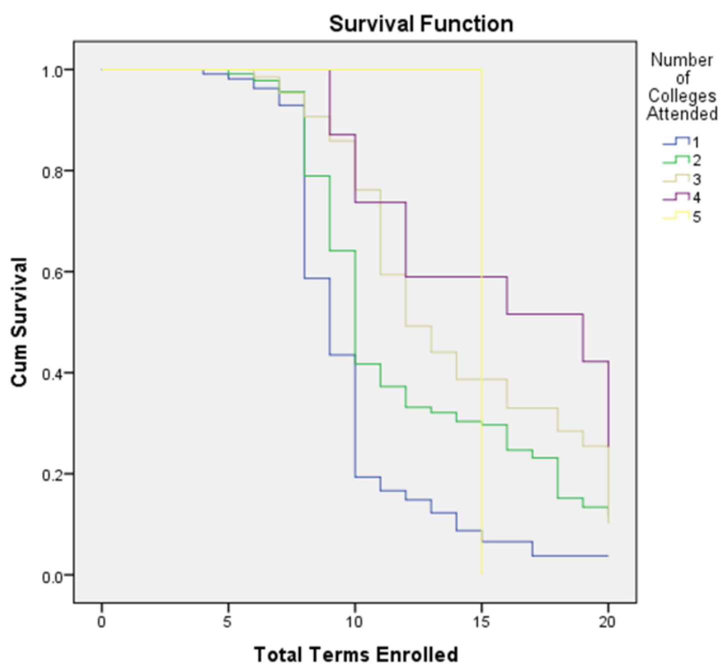

Figure 4.4 *Survival Curve: Number of Colleges Attended*

Table 4.23 provides a pairwise comparison for each level of *number of colleges attended*. This table provides the probability of survival for those who attended only one institution was significantly different from those that attended two, three, or four institutions, but not those that attended five institutions. The survival curve for those that attended five institutions is different from the rest due to the small N for that group.

Table 4.23

Pairwise Comparisons for Number of Colleges

(I) NBR_COLLEGES	(J) NBR_COLLEGES	Wilcoxon (Gehan) Statistic	df	Sig.
1	2	30.741	1	< .001
	3	51.048	1	< .001
	4	22.571	1	< .001
	5	2.409	1	.121
2	1	30.741	1	< .001
	3	12.217	1	< .001
	4	7.824	1	.005
	5	.699	1	.403
3	1	51.048	1	< .001
	2	12.217	1	< .001
	4	1.169	1	.280
	5	.218	1	.640
4	1	22.571	1	< .001
	2	7.824	1	.005
	3	1.169	1	.280
	5	.056	1	.813
5	1	2.409	1	.121
	2	.699	1	.403
	3	.218	1	.640
	4	.056	1	.813

a. Comparisons are exact.

Figure 4.5 displays the survival function in groups by *summer school enrollment*. Those who attended summer school had a survival function with an overall lower probability of survival from one time period to the next. This indicated that students who attend summer school had a higher probability of graduation.

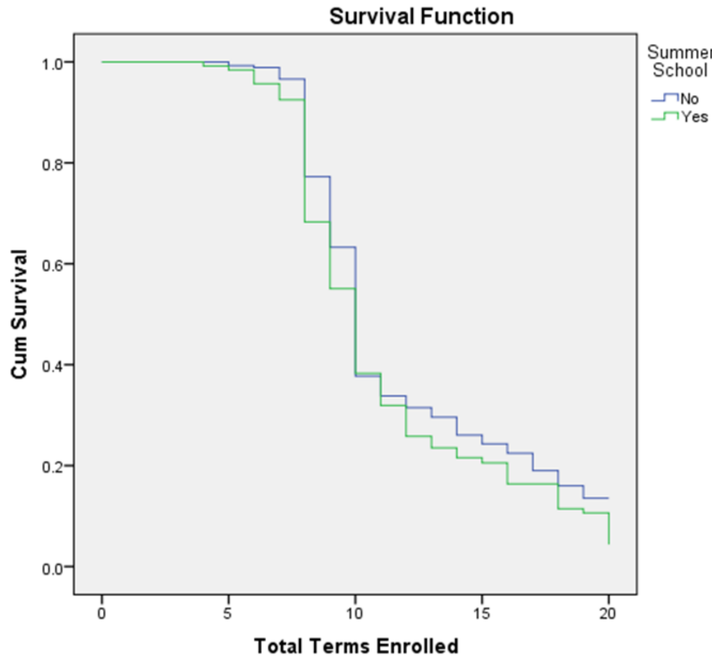

Figure 4.5 *Survival Curve: Summer School Enrollment*

The next step in the analysis was the calculation of the proportional hazards model. This model was calculated using the Cox Regression procedure which models the time to a specific event (*graduation*) based upon simultaneously entered predictor variables. The model produced significantly predicted hazard at χ^2 (13) = 200.535, p < .001. The following variables were found to be significant in predicting survival: *high school GPA*, *number of colleges attended,* and *summer school enrollment* (Table 4.24). Increases in high school GPA and decreases in number of colleges attended resulted in reduced survival time. Summer school enrollment also decreased survival time.

Table 4.24

Survival Analysis Variables

	B	SE	Wald	df	Sig.
ENDER	.055	.104	.280	1	.597
ETHNIC			23.874	6	< .001
HSINSTATE	.144	.152	.902	1	.342
HSGPA	.873	.140	38.817	1	< .001
NBR_COLLEGES	-.505	.080	39.511	1	< .001
FRSTSCHL	-13.515	92.889	.021	1	.884
SUMMER	-.359	.105	11.574	1	< .001
CNT_MLTENRL	.099	.065	2.318	1	.128

The calculated survival function is shown in Figure 4.6. The survival function shows for each term interval the proportional likelihood of surviving to the next term. This survival function takes into account all covariates, and can be used to generally describe the population.

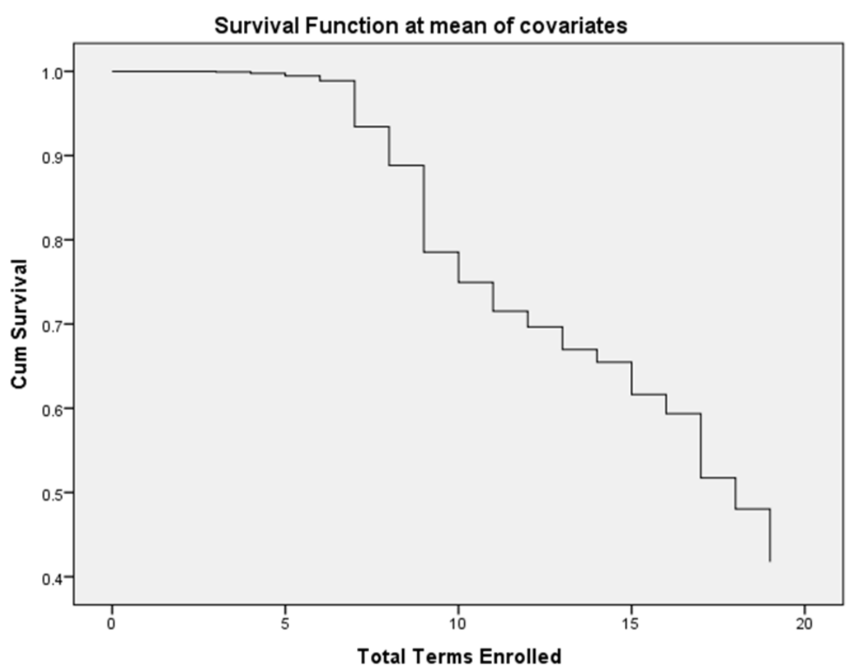

Figure 4.6 *Survival Function*

For this particular data set, it was easier to interpret these data by looking at the hazard function instead of the survival function (Figure 4.7). The hazard function shows the probability of the hazard (in this case graduation) for each subsequent term as predicted for the population. These predictions fall into line with observed values.

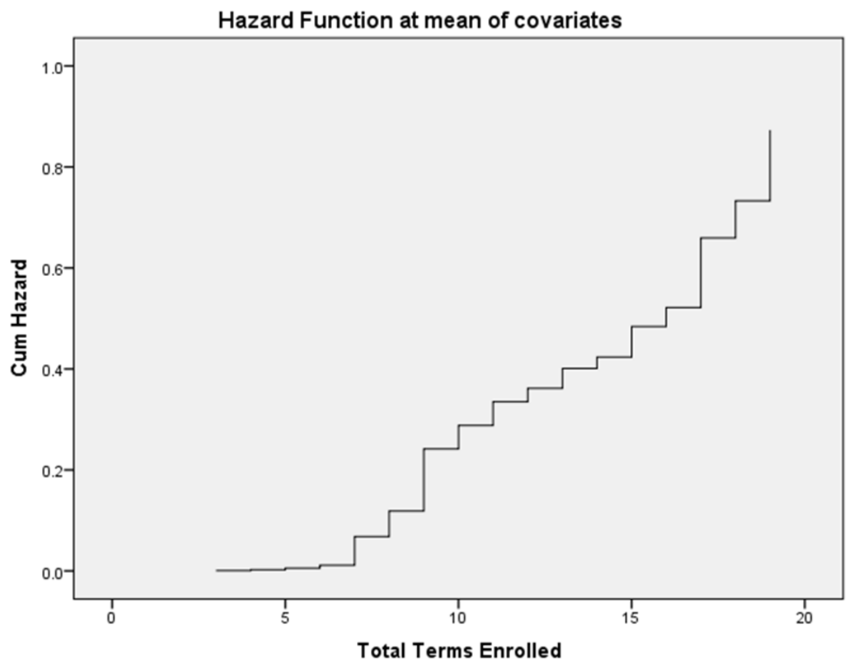

Figure 4.7 *Hazard Function*

Summary

Three multivariate analyses were conducted on a single sample. For each analysis, multinomial logistic regression, discriminant analysis, and survival analysis, models were developed that predicted group membership. All models were able to correctly classify the cases with accuracy at or above 90%.

In each of the models *high school GPA, number of total institutions attended,* and to some extent *summer enrollment* were found to be significant predictor variables (Table 4.25). In both the logistic regression and the discriminant analysis, being enrolled (*ratio* *of semesters enrolled*) was a more significant predictor than the amount of full-time enrollment (*ratio of semesters enrolled full-time*). Additionally, the *total number of institutions attended* had a greater impact on all of the models than the *number of semesters of multi-institutional attendance*. Lastly, both logistic regression and discriminant analysis had poor classification rates for students who were still enrolled at the end of the study.

Table 4.25

Summary of Significant Variables

	Logistic Regression	Discriminant Analysis	Survival Analysis
Significant Variables	• High School GPA • Number of total institutions attended • Ratio of semesters enrolled • Ratio of semesters enrolled full-time • In-state High School	• High School GPA • Number of total institutions attended • Ratio of semesters enrolled • Ratio of semesters enrolled full-time	• High School GPA • Number of total institutions attended • Summer school enrollment • Ethnicity
Classification Rates	94.9%	91.6%	N/A

In the univariate survival analysis, only two variables had significant between group differences in survival times. Students who attended fewer institutions were more likely to graduate earlier than students attending more institutions. Additionally, students who enrolled in summer terms were also more likely to graduate earlier than students who did not enroll in summer terms.

Chapter Five: Discussion

Public universities and state higher education systems are increasingly being held accountable for graduation, persistence and retention rates (Heller, 2011). In chapter two it was established that research on these issues have been dominated by studies at the institutional level. However, the recent focus of research has included studies of longitudinal trends from large multi-institutional data sets (Appendix A). This change was driven by the recognition that student enrollment pathways are increasingly diverse and that these enrollment pathways make it difficult to measure persistence and retention from a single institutional viewpoint. Additionally, researchers are finding that national graduation rates are underreported due to the current federal standards in the definition of 4- and 6-year graduation rates. In response to these perspectives, researchers are seeking more accurate methods of identifying the predictors of student completion. Administrators are utilizing a variety of research findings to develop retention programs at their institutions.

This study applied three different statistical tests (logistic regression, discriminant analysis, and survival analysis) to a random sample of 600 students who graduated from high school in 2000 and subsequently attended the University of Nevada, Reno as a degree-seeking student for at least one semester. The findings demonstrate how sample selection, periods of observation, assumptions and variables could affect the utility and generalizability of this type of research. While the outcome of each statistic individually can inform practice at individual institutions, it is the information provided by each statistic in comparison with the others that yields the most information.

Review of Outcomes

Logistic regression. The first statistical analysis was a multinomial logistic regression. Two different models were obtained from this analysis, one that included summer in the enrollment progression variables and one that did not. While both models classified the subjects with an accuracy rate of above 90% and both models were significant at $p < .001$, the model that included summer had a higher classification rate 95.4% versus 94.5%. The other important finding in the regression models was the effect of multi-institutional attendance. Nearly 25% of the sample had attended more than one institution in the same semester. However, the number of semesters a student was enrolled in multiple institutions had no significant effect on the likelihood of graduation. What did have a significant effect was the *number* of total institutions a student attended over the observation period. According to the models, for each additional school a student attends the likelihood of graduation decreases. Beta values for the *ratio of total enrolled semesters* was larger than the beta values for the *ratio of total full-time semesters*. This has several implications for administrators at institutions of higher education as they try to increase retention. Simply stated, student persistence, even at the smallest measurable level (i.e. taking just one class) significantly increases the likelihood of graduation. It is therefore preferable to find ways to encourage a student to remain enrolled, even at a part-time status, than allowing the student to drop-out for any period of time.

Discriminant analysis. The discriminant analysis model produced similar results. While several variables were found to be significant contributors to the model, only two had sufficiently low Wilks' Lambda values to be important in interpreting the results, *ratio of regular enrolled semesters* and *ratio of regular full-time enrolled semesters.* This closely aligned with the results from the logistic regression and can be interpreted in a similar manner, in that student enrollment term after term is the best predictor of graduation.

With both models, there were challenges in correctly classifying those students still enrolled at the end of the observation period. The summer logistic regression model only classified 26% of the cases correctly and the discriminant analysis only 19%. This indicates that enrollment data alone cannot predict who will still be enrolled at the end of a ten-year observation period and that other, unknown variables are at play.

Survival analysis. Survival analysis is not designed to identify predictors for graduation. Instead it is designed to identify those predictors that are associated with students who have a *shorter* time to graduation. This statistic can be highly useful to help identify those characteristics that can increase the hazard of a measured event, in this study graduation. The results indicate that the number of schools attended will decrease the hazard of graduation while attending summer school can increase the hazard. In this study a higher hazard of graduation is preferable.

Comparison of Statistical Analyses

In combining the results from all of the statistical methods, the following observations were consistent. First, each new school students attend will decrease their probability of graduation and will increase their time to graduation. Second, enrollment, at all levels (full- or part-time), will increase the probability of graduation and will decrease the time to graduation. Lastly, multi-institutional enrollment, in and of itself, does not have a significant impact on graduation. This is mitigated by the fact that in order to have multi-institutional enrollment a student needs to attend a second institution which does have a significant impact. This could be interpreted as enrollment in more than one institution could be a sign of enrollment intensity or persistence, which has the greatest positive effect on graduation outcomes.

Sample selection. An important consideration in designing a study targeting alternative enrollment pathways is in the selection of the sample used. The intention of the study was to look at a broader view of student persistence, including students on a variety of different enrollment pathways, particularly those on pathways leading to multi-institutional attendance. Many studies of graduation in higher education focus on the first-time, full-time cohort. However, this study found that the type of sample can limit the research to answering questions that only relate to institutional retention rather than student completion. The exclusion of students on alternative enrollment pathways results in under reporting of completions. This under reporting can have negative impacts on both funding and public perceptions of higher education.

In order to obtain a broader view, researchers must expand the definition of their cohorts to include students on alternative enrollment pathways, including those engaging in multi-institutional attendance. Adelman (2004, 2006) in his national longitudinal studies used cohorts as defined by high school graduating classes rather than single institutional data. This type of sample selection allows the researcher to include cases where the student starts late, transfers in, and even transfers out. Including alternative student enrollment pathways often increases completion rates. For example, the institutional 10-year graduation rate for the Fall 2000 admissions class at the primary institution in this study was 57%. However, when a regional perspective was used for the sample (high school graduation class and graduation at any institution) there was a 76.5% graduation rate over the same period of observation. Additionally, 5% of the students were still enrolled at the end of the study.

This revised method of sampling can bring challenges. In order to obtain a complete picture of a student's pathway, institutional researchers must use national databases such as provided by the National Student Clearinghouse. While the Clearinghouse contains a more robust picture of student enrollment, numerous data elements such as major, specific class enrollment patterns and grades are neither consistently available nor accurate. Many institutional researchers are challenged to forgo the depth of information they may access through their institutional datasets to look at a more regional/national perspective. State systems may have access to multi-institutional data sets, but may be challenged in coordinating any strategic interventions or regional changes among numerous institutions. Even with these challenges however, this study used primarily only data obtained through the Clearinghouse and was able to obtain model classification rates of over 90%.

Periods of observation. Traditional models of research in higher education have typically only considered an observation period of four to six years stemming from the cohort defined by the Student Right to Know Act. Over the last decade however there has been an increase in the number of longitudinal studies exceeding these time periods (See Appendix A). The data from this study supported the utilization of a longer time frame. Even with a ten-year observation period, nearly 5% of the sample was still enrolled at the end of the ten-year period, still pursuing their first baccalaureate.

Assumptions. One of the most significant challenges to the use of more sophisticated statistics is in the number and type of assumptions that must be met about the data before the ability of the statistic to find significant differences starts to diminish. The sample for this data study intentionally included students on as many different educational pathways as possible. While removing outliers was a relatively straightforward process, the requirements of normality, linearity and homogeneity of variance were much more difficult to meet. It is the nature of these data to include skewed variables, especially if they represent students on more unique pathways such as multi-institutional attendees. This can restrict researchers from using statistics that may be considered more rigorous if the intent is to ensure that as many cases as possible are included.

For this study, a discriminant analysis was conducted, but there was a concern of the accuracy of the results. The data did not meet the assumption of homogeneity of variance, generally required before running this statistic. While Tabachnick and Fidell (1996) stated that discriminant analysis is robust to the lack of homogeneity of variance as long as the variance is due to skewed data rather than outliers, there were still problems interpreting the overall results. While several of the variables were found to be significant contributors to the overall model (Table 4.19) many of those had high Wilks' Lambda values indicating that they explained relatively little of the variance within the actual data set. The similarities in comparison with the logistic regression however increase confidence in the results of the discriminant analysis.

Variables used. Overall, the dataset used was simple, term based enrollment data from the National Student Clearinghouse combined with gender, ethnicity, and high school GPA from the primary institution's student information system. Even with this limited enrollment data, classification rates for the models created using logistic regression and discriminant analysis were above 90%. Additionally, both logistic regression models had a high Nagelkerke R-Square value (both cases above 0.9), meaning they both explain over 90% of the variation observed within the sample. This is important as it implies that little above and beyond simple enrollment behaviors and limited admissions/high school data is needed to predict graduation.

One of the biggest challenges to using statistics such as logistic regression or discriminant analysis is transforming data that includes several observations over a period of time for each case into a variable that can be used within the analysis. For this study, the decision was made to create several enrollment intensity variables out of the raw data. This approach was chosen as the new variables were accurate representations of how much and how intense (i.e. how often they were enrolled full-time or how often they were enrolled as a multi-institutional student) were the observed enrollment pathways.

Implications for Practice

Based on chapter two, there is a plethora of research available to administrators regarding persistence and retention. However, many researchers still use the concepts of retention and persistence interchangeably. As noted in chapter two, there is a greater need to differentiate between *student* persistence and *institutional* retention. With the pressure for states to demonstrate successful efforts to improve graduation and completion rates, such a differentiation can matter. As administrators review the research it is important to also review the sample selection procedure used to obtain those findings to determine if the results are applicable to students on alternative enrollment pathways or if they only apply to traditionally enrolled students.

Studies that include alternative enrollment perspectives on student persistence also get a much clearer and *positive* picture of the success on graduating students in higher education. For example, the institutional 10-year graduation rate for the Fall 2000 admissions class at the primary institution in this study is 57%. However, when a multi-institutional perspective was used for this sample (high school graduation class and graduation at any institution) there was a 76.5% graduation rate. The argument that can be made is that if we, as national public policy, follow student persistence in addition to institutional retention, we can gain a much more accurate picture of postsecondary completion rates and alternative enrollment pathways to better evaluate the outcomes of higher education. A second question that should be asked is whether the interventions and programs being developed and funded by the federal government, state and institutional entities to improve college going and graduation rates are focusing on areas that may not be sufficiently defined. If more students are successful than the current data predicted, then where exactly are the points in the student enrollment pathways where drop-out and stop-out really occur? Additionally, which of the enrollment pathways are natural consequences of student lives and which fall under institutional pressure or influence?

Lastly, the best predictor of graduation is persistence. For each semester a student is enrolled, whether full-time or at more than one institution, the likelihood of graduation increases, and the time to graduation decreases. This creates an important policy question for administrators. Many policies are currently set up to encourage students to be full-time or to be only enrolled in one institution. If a student's only choices are between being enrolled full-time or dropping out, this may lead to increased dropout rates; whereas if an institution encourages student persistence, even at a part-time level, it may actually help the institutional graduation rates over a longer time frame. This assistance could take the form of institutional aid to part-time students, setting up more formalized co-admission programs with other regional institutions that make it easier for a student to attend multiple institutions at the same time if doing so increases the student's likelihood of remaining enrolled.

Recommendations

Ultimately, it is not the actual selection of the statistic that can affect outcomes, but how the parameters of sample selection and periods of observation are chosen. In order to assure that a study includes outcomes that can be applied to alternative enrollment pathways, care must be taken to ensure all possible enrollment pathways are included in the overall study. The results of this study suggest changes to current federal and state accountability policies as well as to institutional retention practices. Many of the suggestions included here could help increase student persistence; however, they may result in poorer metrics at the institutional level. Institutions are often measured, and even funded, on the number of full-time students they have attending, so increasing the number of part-time students could have short term undesired effects.

Federal, state and institutional governing boards should reconsider the current metrics institutions are measured on, such as retention and graduation rates of first-time full-time cohorts over a 4- to 6-year period. These groups should also look at policy recommendations to develop complimentary metrics that describe enrollment patterns and graduation rates in a more comprehensive manner to take into account students who swirl. In order to gain a fuller understanding of graduation and completion at the state and regional level, boards and legislators should consider finding new ways to revise metrics that measure and track students.

The students who were still enrolled after ten years defied classification attempts. These students are still undefined and unknown. Institutions may argue that this is an undesirable outcome, as the prevailing belief is that graduation needs to occur in a timely manner, but if these students are still on their way to a degree, then it is important to research and understand these sub-populations. Institutions also need to pay attention to students who do not return. Are they stop-outs, drop-outs, or transfers? Students who ultimately go on to obtain a degree should still be reflected as part of an institution's completion rates, even if the metric is new and different from current measures.

Lastly, there is a lack of understanding in why students choose the enrollment pathways that they do. A question that needs to be explored is: Are student's choosing these pathways intentionally; do these pathways coincide with student life choices? Or are these pathways the result of inadequate sources of funding or failure to realize satisfactory academic progress? These reasons have different implications in how institutions and/or governing boards respond. In order to understand these choices further, qualitative research must be conducted.

Conclusion

This study compared statistical methods and redefined sample selection and observation periods in order to obtain a fuller understanding of college student persistence and completion rates. If graduation rates continue to be utilized to fund and publically rank the success of institutions, then the definitions and methods used to derive these data are important. Public policy makers and institutional leaders utilize this data to direct resources and improve performance. Students who have unconventional enrollment pathways need to be included in our metrics of student success and persistence. They are an overlooked and oft excluded population who are finding ways to succeed even without the full attention of policy makers and administrators. What has been regarded as the exception is on the way to becoming the norm. In this 10-year analysis nearly a quarter of the students attended more than one institution in the same semester. There are signs that this phenomenon continues to increase, particularly with the advancement of online learning and MOOC's (massive open online course). If we desire to improve college and university graduation rates, then it is imperative that we know what our rates actually are. We need to know about the alternative enrollment pathways and completion rate of all of our students and not just the ones who attend in the previously defined traditional manner of first-time, full-time.

References

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment. Washington, DC: U.S. Department of Education.

Adelman, C., Daniel, B., and Berkovits, I (2000). Postsecondary attainment, attendance, curriculum, and performance: Selected results from the NELS:88/2000 Postsecondary Education Transcript Study (PETS), 2000. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Adelman, C. (2003). The story-lines of multi-institutional attendance. Paper presented at the Association for Institutional Research Forum.

Adelman, C. (2004). Principle indicators of student academic histories in postsecondary education, 1972-2000. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Adelman, C. (2005). Moving into town—and moving on: The community college in the lives of traditional-age students. Washington, DC: U.S. Department of Education.

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

Adelman, C. (2006b, October 13). The propaganda of numbers. *Chronicle of Higher Education*.

Archibald, R.B., and Feldman, D.H. (2008). Graduation rates and accountability: Regressions versus production frontiers. *Research in Higher Education, 49*(1), 80-100, DOI: 10.1007/s11162-007-9063-6.

Astin, A.W. (1993). What matters in college? Four critical years revisited. San Fransisco: John Wiley & Sons.

Astin, A.W. (2005-2006). Making sense out of degree completion rates. *Journal of College Student Retention, 7*(1-2), 5-17.

Bach, S.K., Banks, M.T., Kinnick, M.K., Ricks, M.F., Stoering, J.M., and Walleri, R.D. (2000). Student attendance patterns and performance in an urban postsecondary environment. *Research in Higher Education, 41*(3), 315-330.

Bahr, P.R. (2009). College hopping: Exploring the occurrence, frequency, and consequences of lateral transfer. *Community College Review, 36*(4), 271-298, DOI: 10.1177/0091552108330903.

Barkley, S.M. (1993). A synthesis of recent literature on articulation and transfer. *Community College Review, 20*(4), 38-50, DOI: 10.1177/009155219302000405.

Bearer-Friend, J.. (2009, November). Measuring student achievement at postsecondary institutions. *NGA Center for Best Practices Issue Brief*, http://www.epi.soe.vt.edu/perspectives/policy_news/pdf/0911MEASURING.pdf.

Borden, V.M.H. (2004). Accommodating student swirl: When traditional students are no longer the tradition. *Change, 36*(2), 10-17.

Cabrera, A.F. (1994). Logistic regression analysis in higher education: An Applied Perspective. In Smart, J.C. (ed.) *Higher Education: Handbook of Theory and Research, Volume X*, New York: Agathon Press, 225-256.

Calcagno, J.C., Crosta, P., Bailey, T., & Jenkins, D. (2007). Stepping stones to a degree: The impact of enrollment pathways and milestones on community college student outcomes. *Research in Higher Education, 48*(7), 775-801, DOI: 10.1007/s11162-007-9053-8.

Chimka, J.R., Reed-Rhoads, T., & Barker, K. (2007-2008). Proportional hazards models of graduation. *Journal of College Student Retention, 9*(2), 221-232.

Cohen, H. & Ibrahim, N. (2008). A new accountability metric for a new time: A proposed graduation efficiency measure. *Change 40*(3).

Curtin, T.R., Ingels, S.J., Wu, S., Heuer, R. (2002). *User's manual: NELS:88 Base-year to fourth follow-up: Student component data file.* (NCES 2002323). Retrieved from the U.S. Department of Education, Center for Education Statistics website: http://nces.ed.gov/pubs2002/2002323.pdf.

de los Santos, A.G. and Wright, I. (1990). Maricopa's swirling students: Earning one-third of Arizona State's bachelor's degrees. *Community, Technical, and Junior College Journal, 60*(6), 32-34.

Student Right-To-Know Act, 34 C.F.R. pt. 668 (1996).

Goldrick-Rab, S. (2006). Following their every move: An investigation of social-class differences in college pathways. *Sociology of Education, 79*(1), 61-79.

Grubb, W.N. (1991). The decline of community college transfer rates. *Journal of Higher Education, 62*, 194-217.

Hagedorn, L.S. (2006). Traveling successfully on the community college pathway. Los Angeles, CA: Transfer and Retention of Urban Community College Students (TRUCCS)

Heinze, M.C. and Daniels, J.L. (1970). The transfer of students into community colleges. (ERIC Document Reproduction Service No. Ed 050 723).

Jones-White, D.R., Radcliffe, P.M., Huesman Jr., R.L., & Kellogg, J.P. (2009). Redefining student success: Applying different multinomial regression techniques for the study of student graduation across institutions of higher education. *Research in Higher Education, 51(*2), 154-174, DOI: 10.1007/s11162-009-9149-4.

Kuh, G.D., Kinzie, J., Schuh, J.H., Whitt, E.J., & Associates (2005). *Student success in college: Creating conditions that matter*. San Francisco: Jossey-Bass.

Kuh, G.D. (2009). What student affairs professionals need to know about student engagement. *Journal of College Student Development, 50*(6), 683-706.

Li, D. (2010). They need help: Transfer students from four-year to four-year institutions. *The Review of Higher Education, 22*(2), 207-238.

Marti, C.N. (2008). Latent postsecondary persistence pathways: Educational pathways in American two-year colleges. *Research in Higher Education, 49*(4), 317-336, DOI: 10.1007/s11162-007-9083-2.

McCormack, A.C. (1997). Transfer behavior among beginning postsecondary students: 1989-94. Washington, DC: National Center for Education Statistics.

McCormack, A.C. (2003). Swirling and double-dipping: New patterns of student attendance and their implications for higher education. *New Directions for Higher Education, 121*, 13-24.

Mertler, C.A. & Vannatta, R.A. (2005). Advanced and Multivariate Statistical Methods: Practical Application and Interpretation, 3rd Ed. Glendale, CA: Pyrczak Publishing, 348 pgs.

Murtaugh, P.A., Burns, L.D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education, 40*(3), 355-371.

National Student Clearinghouse Research Center (2011). Snapshot Report. http://research.studentclearinghouse.org. Retrieved December 10, 2011.

National Student Clearinghouse. (2013). *Enrollment reporting: Programming and testing guide.* Retrieved from http://www.nationalstudentclearinghouse.com/colleges/files/EnrollRept_ProgrammingandTestingGuide.pdf

Pascarella, E.T. & Terenzini, P.T. (2005). *How college affects students: A third decade of research.* San Fransisco: Jossey-Bass.

Peng, C.J., So, T.H., Stage, F.K., & St. John, E.P. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999. *Research in Higher Education, 43*(3), 259-293.

Peter, K. and Cataldi, E.F. (2005). The road less traveled? Students who enroll in multiple institutions. (NCES 2005-157). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Reason, R.D. (2009). An examination of persistence research through the lens of comprehensive conceptual framework. *Journal of College Student Development, 50*(6), 659-682.

Ronco, S.L. (1995). How enrollment ends: Analyzing the correlates of student graduation, transfer, and dropout with a competing risks model. Paper presented at the 35th Annual Forum of the Association for Institutional Research. Boston, MA, May 1995.

Roska, J. (2009). Bachelor's degree completion across state contexts: Does the distribution of enrollments make a difference? *Research in Higher Education, 51*(1), 1-20, DOI: 10.1007/s11162-009-9146-7.

Russell, A. (2009). Update on graduation rate reporting: Issues and opportunities. *American Association of State College and Universities, A Higher Education Policy Brief*, http://www.aascu.org/policy/publications/policy-matters/2009/graduationrate.pdf

Scott, M., Bailey, T., & Kienzl, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the U.S. *Research in Higher Education, 47*(3), 249-279, DOI: 10.1007/s11162-005-9388-y.

Seabring, P., Tourangeau, R., Spencer, B., Glusberg, M., Campbell, B., and Singleton, M. (1987). *The National Longitudinal Study of the High School Class of 1972 (NLS-72), Fifth follow-up (1986): Final technical report* (NCES 87100). Retrieved from the U.S. Department of Education, Center for Education Statistics website: http://nces.ed.gov/pubs87/87100.pdf.

Shaprio, D., Dundar, A., Chen, J., Ziskin, M., Park, E., Torres, V., and Chiang, Y. (2013). Completing college: A state-level view of student attainment rates. *Signature Report.* Retrieved from the National Student Clearinghouse Research Center website: http://www.studentclearinghouse.info/signature/4state/

Tabachnick, B.G. & Fidell, L.S. (2007). *Using Multivariate Statistics, 5th Ed.* Boston: Pearson. 963 pgs.

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

Tinto, V. (2006-2007). Research and practice of student retention: What next? *Journal of College Student Retention, 8*(1), 1-19.

Townsend, B.K. (2000). Rationales of community colleges for enrolling reverse transfer students: A second chance for whom? *Community College Journal of Research and Practice, 24*(X), 301-311.

Townsend, B.K. (2001). Redefining the community college transfer mission. *Community College Review, 29*(2), 29-42, DOI: 10.1177/009155210102900203.

Townsend, B.K. and Dever, J.T. (1999). What do we know about reverse transfer students? *New Directions for Community Colleges, 106*, 5-14.

Visser, M.S., Luwel, M., & Moed, H.F. (2007). The attainment of doctoral degrees at Flemish universities: A survival analysis. *Higher Education, 54*(5), 741-757. DOI: 10.1007/s10734-006-9021-9

Willet, J.B. & Singer, J.D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research, 61*(4), 407-450.

Willet, J.B. & Singer, J.D. (1993). It's about time: Using discrete-time survival analsysis to study duration and the timing of events. *Review of Educational Research, 18*(2), 155-195.

Willet, J.B. & Singer, J.D. (1995). It's déjà vu all over again: Using multiple-spell discrete-time survival analysis. *Journal of Educational and Behavioral Statistics, 20*(1), 41-67.

Winter, P.A. & Harris, M.R. (1999). Community college reverse transfer students: A field of study of a nontraditional student group. *Community College Review, 27,* 13-29.

Winter, P.A., Harris, M.R., and Ziegler, C.H. (2001). Community college reverse transfer students: A multivariate analysis. *Community College Journal of Research and Practice, 25*, 271-282.

Wolf-Wendel, L., Ward, K., & Kinzie, J. (2009). A tangled web of terms: The overlap and unique contribution of involvement, engagement, and integration to understanding college student success. *Journal of College Student Development, 50*(4), 407-428.

Zahs, D., Pedlow, S., Morrissey, M., Marnell, P., and Nichols, B. (1995). *High school and beyond fourth follow-up methodology report* (NCES 95426). Retrieved from the U.S. Department of Education, Center for Education Statistics website: http://nces.ed.gov/pubs95/95426.pdf.

Appendix A: Studies of Enrollment Pathways

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Adelman	1999	Degree completion	Demographic, high school enrollment data, college enrollment data	High School and Beyond/Sophomore Cohort 1988	N	8,873	13 years	Descriptive analysis, logistic regression	Does not analyze concurrent enrollment
Adelman	2003	Attendance pattern	Transcript data	National Education Longitudinal Study of 1988 (NELS:88/2000)	N	8,887	12 years	Descriptive analysis	
Adelman	2004	Degree completion	Demographic, high school background, postsecondary entry, first-year postsecondary performance, financial aid, attendance patterns, extended postsecondary performance, final factors	National Longitudinal Study of the High School Class of 1972 (NLS:72/86) High School & Beyond Longitudinal Study of 1980 Sophomores (HS&B-So:80/92) National Education Longitudinal Study of 1988 (NELS:88/2000)	N	Various	Various	Descriptive analysis, regression	Does not analyze concurrent enrollment

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Adelman	2005	Transfer	Demographic and transcript data	National Education Longitudinal Survey: 88/2000	N	8,900	12 years	Descriptive analysis, regression	Does not analyze concurrent enrollment
Adelman	2006	Degree completion	Demographic, high school background, postsecondary entry, first-year postsecondary performance, financial aid, attendance patterns, extended postsecondary performance, final factors	National Education Longitudinal Survey: 88/2000 1992 12th graders who subsequently attended a four-year college at any time through Dec 200, who earned their high school diploma by December 1996, who presented complete high school transcript records and a senior year test score, whose postsecondary records were complete, and whose socioeconomic status was known	N	Not availab le	12 years	Descriptive analysis, regression	Does not analyze concurrent enrollment

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Adelman, Tabs, Berkovits	2000	N/A	Degree received, academic intensity, major, credits, courses details, institution(s), GPA, SES	National Education Longitudinal Survey: 88/2000	N	12,100	12 years	Descriptive analysis	Data was recorded in such a manner as to track how many institutions, but not necessarily what credit taken could be assigned to each institution
Alfonso	2006	Baccalaureate degree attainment	Individual specific characteristics, educational pathway, institution of first attendance, student's educational expectations	National Education Longitudinal Survey 1988	N	12,144		Structural Equation Model	The observation period for each student stopped after the first break in enrollment

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Bach, Banks, Kinnick, Ricks, Stoering, and Walleri	2000	Graduation	Demographics, transcript data, and transfer evaluation data	Undergraduates at their first university enrollment and whose last enrollment at a community college was during the 1990-1991 academic year	N	336	Not clear	Descriptive, ANOVA	Did not track concurrent community college enrollment, but did examine concurrent university/community college enrollment; participants must have transferred from community college to university
Bahr	2009	Lateral transfer and completion of credential	Sex, age at college entry, race/ethnicity, receipt of financial aid, and duration of attendance	Fall 1995 cohort of first-time college students who enrolled in any of California's 106 semester-based community colleges	Y	156,188	6 years	Descriptive analysis, regression	For each observed semester, each student was assigned to the college in which he or she enrolled in the greatest number of courses; only counted the highest credential earned, the first time it was earned

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
del los Santos and Wright	1990	N/A	Transfer and graduation	Studens attending both Arizona State University and Maricopa Community College District from 1982-1989	N	-	8 years	Descriptive	
Goldrick-Rab	2006	Enrollment Pathway	Demographics and high school preparation	National Education Longitudinal Survey 2000, students who had participated in all follow ups, attended at least one postsecondary institution, had a complete transcript record, and began at a four-year institution	N	4,628	12 years	Multinomal logistic regression	Only looked at students who began their education at a four-year institution

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Kirk-Kuwaye and Kirk-Kuwaye	2007	Engagement Patterns	Previous institution type, participation in new student orientation, ethnicity, gender	Fall 2005 incoming transfer students who had only previously attended one institution, had between 24-88 transfer credits, and were from non-science and non-professional majors	N	17	N/A	Interviews	Limited sample size drawn from a very specific population
Li	2010	Baccalaureate degree attainment within 6 years	Student pathways: (1) stayed, (2) stop out, (3) interrupted transfer, and (4) continuous transfer Student characteristics and family background	Beginning Postsecondary Students Longitudinal Study 1995-1996, first-time degree-seeking traditional students who originally attended a four-year institution	Y	2,990	6 years	Descriptive analysis, regression	Only looked at first-time students at four-year institutions; doesn’t look at students who had not graduated or were not still continuously enrolled

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Marti	2007	Attendance patterns	Enrollment and results from Community College Student Report (engagement)	Stratified random sample of respondents to the Community College Student Report from Florida	--	--	3 years	Latent Trajectory Analysis and Logistic Regression	Single institution data and limited observation window
McCormick	1997	Certificate or degree attainment	Student characteristics; student goals; student satisfaction, institutional data, academic information	Beginning Postsecondary Students Longitudinal Study, BPS:90/94	Y	Not available	5 years	Descriptive analysis, t-tests, multiple linear regression	Only examined first-time students; only investigated those multi-institutional attendees that were defined as transfer
Peter and Cataldi	2005	Persistence, Attainment, and Time to Degree	Enrollment and demographic data	Beginning Postsecondary Students Longitudinal Study 1996/01 and Baccalaureate and Beyond Longitudinal Student 2000/01	N	10,000 +	6+ years	Descriptive, t-tests, two-way ANOVA, multiple linear regression	Only reviewed a subset of multi-institutional pathways

Author(s)	Year	Dependent Variable(s)	Independent Variable(s)	Sample	FTFT	N	Length	Method	Limitation(s)
Winter, Harris and Ziegler	2001	Reverse transfer status (completer or non-completer)	Survey including personal data, 17 questions about reasons and 6 about goals for attending a community college	Kentucky reverse transfer students enrolled at 14 different community colleges	N	885	N/A	Two group, stepwise discriminant analysis	Limited to a specific group of multi-institutional attendees