2016

Type of First Term Course Failure and Community College Degree Completion

Jill Stearns
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

Jill Stearns

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Alice Eichholz, Committee Chairperson, Education Faculty
Dr. Daniel Salter, Committee Member, Education Faculty
Dr. Gerald Giraud, University Reviewer, Education Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2015

Abstract

Type of First Term Course Failure and Community College Degree Completion

by

Jill Stearns

MS, National University, 2002

BS, California Polytechnic State University San Luis Obispo, 1990

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

February 2016

Abstract

Community colleges are the largest segment of higher education institutions in the United States providing access to historically underserved populations and growing numbers of first generation college students. Increasing college degree attainment is a national priority with new expectations of accountability. Despite decades of educational research, community colleges have startling low completion rates. Within the framework of Tinto's theory of retention, a predictive analytics model could provide community colleges the opportunity to drive custom intervention and support services to students. The purpose of this study was to explore the utility of Biglan's taxonomy for categorizing courses for potential use in a data analytics model to identify students at risk of failure to complete. The quantitative census study used archival data from 1,759 students. Log-linear analysis was used to test the key research question as to whether there is a predictive relationship between type of course failed, as cross-categorized by the dimensions in Biglan's taxonomy, in the first term and failure to complete a degree or certificate within 6 years. The analysis showed that a more parsimonious model, based on the interaction term for the life/nonlife and pure/applied Biglan categories, appeared related to completion, although no standardized residual was significant. A larger and more diverse sample may be necessary to determine the true effectiveness of Biglan's taxonomy as a classification schema in a predictive analytics model of degree completion. Based on these results, first term course failure appears to be a logical point for programmatic support that could lead to higher levels of associate degree completion opening doors of employment opportunity through education, thus supporting social change.

Type of First Term Course Failure and Community College Degree Completion

by

Jill Stearns

MS, National University, 2002

BS, California Polytechnic State University San Luis Obispo, 1990

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

February 2016

Dedication

In dedication to the four generations of women in education I am blessed to call family. My maternal great grandmother, Gertrude Preston Hawthorne, who taught in a one-room school house in Missouri. My maternal grandmother, Vula Cottengim Bowers, who always led from the side. My amazing mother, Shirley Bowers Hodges, who lived her life in service to the students of Avenal High School. And my beloved sister, Janet Hodges Rodrigues, who makes it all look easy. Because of your example, I found my professional home in education.

Acknowledgements

Keith, my precious husband and faithful friend, I cannot thank you enough for your patience, practicality, and humor. I can't wait to see what life looks like without a paper requiring my attention. To Cameron and Sam, who will always be my boys, thank you for the reminder of what is really important. I love you all and am amazed by what you have become.

Thank you to my mentors, formal and friendly, who have encouraged me on this long path. Dr. Frank Gornick, Dr. Joan Smith, Dr. Willard Lewallen, and Dr. Mary Retterer, your support was invaluable and inevitably just in time!

Sincerest thanks to my dissertation committee who provided the expertise and critical feedback I needed. I appreciate the amazingly rapid response time you demonstrated throughout the process.

Dad, thank you for always believing I could get this done. I'm looking forward to our next adventure.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Degree completion in community colleges gained attention in California and nationally in the era of increased accountability of public fund expenditures and transparency of effectiveness (Hickman, 2011; Summers, 2003; Tinto, 2004). Since early research in student attrition and persistence began, community colleges have sought solutions to increase degree completion: for example, Bean (1979) Cabrera, Nora, and Casteñada (1991); Elkins, Braxton, and James (1998); and Tinto (1987) have all conducted studies of the factors of retention. However, low rates of community college degree completion remain a challenge. The challenge now includes serving increasing numbers of first generation students and millennials, two groups of students with characteristics identified as being at risk for attrition (Engle & Tinto, 2008; Tinto, 2004). A significant ongoing issue in community college education is failure to complete within 6 years (Tinto, 2004). This study was designed to explore the opportunity to direct resources to students with characteristics and factors predictive of failure to complete a degree or certificate within 6 years. In particular, the goal was to determine the plausibility of applying predictive analytics to increase retention. This study sought to determine if there was a predictive relationship between the type of course failed in the first term to failure to complete a degree in 6 years.

Social media companies and online retailers employ the science of predictive analytics to deliver customized content to individual users and consumers (Johnson, 2012). In higher education, institutions store vast amounts of student information that may be useful in reducing attrition in a model similar to that employed by Google and

Amazon (Smith, Lange, & Huston, 2012). Application of effective predictive analytics in the community college requires identification of student information data elements with a relationship with completion, in a way that allows analysis of diverse data across institutions. This study examined one possible element in the form of type of course failed in the first term.

This chapter includes a brief overview of the research literature in the areas of student retention, millennials, first generation students, and classification of academic disciplines. The research problem is introduced and the purpose of the study is presented with definitions and assumptions to increase understanding. The theoretical foundation and nature of the study are examined including the scope and limitations of the study. Finally, the significance of the study is presented.

Background

Community colleges serve approximately 45% of students enrolled in higher education institutions (Giancola & Davidson, 2015). The first time students enrolling in community colleges are more likely to be students of color, the first generation of their family to attend college, and have lower incomes as compared to first time students enrolling in other types of colleges and universities (Engle & Tinto, 2008). Further, students enrolling in community colleges work while attending colleges at rates higher than those students attending other types of institutions (Engle & Tinto). Low-income first generation students were nearly four times more likely to leave higher education after the first year than their peers without these risk factors (Engle & Tinto).

Decades of research have assisted in identifying a variety of risk factors of student attrition. The risk factors include: limited academic and social integration (Bean, 1980; Bean & Metzner, 1985; Burns, 2010; Cabrera, Nora, & Castañeda, 1993; Spady, 1971; Tinto, 1975, 1987, 1997, 2004, 2008), inadequate student support services (Engstrom & Tinto, 2008; Murray, 2010; Tinto, 2008), academic preparedness and performance (Elkins, Braxton, & James, 1998; Kalsner, 1991; Summers, 2003), first generation students (Oldfied, 2007), and millennial generation students (Becker, 2009; Elam, Stratton, & Gibson, 2007; Jones & Healing, 2010; Scott-Clayton, 2011; Varallo, 2008). These studies have led the way to the current design of student support services and interventions to increase student completion; however, the challenge of degree completion within 6 years, a national benchmark, persists.

Current state and national reports have indicated that although community colleges have successfully addressed the goal of access to higher education for all who desire to attend, only 11% of low-income, first generation students had earned bachelor's degrees compared to 55% of their more advantaged peers attaining bachelor's degrees (Board of Governors of the California Community Colleges, 2013; Engle & Tinto, 2008). Students who began at community colleges and transferred to 4-year institutions reflected similar disparities in achievement (Giancola & Davidson, 2015). Students who were not low income, first generation students transferred and completed bachelor's degrees at a rate of 24%, while only 5% of low income, first generation students successfully completed bachelor's degrees (Engle & Tinto, 2008, p.2). Despite the volume of research on the risk factors of attrition and effective practices to increase student retention, the

completion rates at community colleges remain low. In California, which has the largest student enrollments in community colleges, only 59% of the first time students who began in fall 2013 persisted into fall 2014 ("First-generation…", 2014).

The California Community Colleges Chancellor's Office (CCCCO) has renewed the focus on student success and completion through a legislated redesign of student support services replacing what was previously termed "matriculation" (Board of Governors of the California Community Colleges, 2013). In the California community college system, matriculation previously referred to the funding stream that compensated colleges for delivery of student services according to the Matriculation Act of 1986, later renamed the Seymour-Campbell Student Success Act of 2012 by the legislature (Board of Governors of the California Community Colleges, 2013). The Student Success Act included the development of an informed education goal, development of educational plans, and declaration of a course of study; elements which had not been included as part of matriculation (Board of Governors of the California Community Colleges). The redesign was intended to align student support services toward degree completion based upon institutional data and research-based practices.

The matter of student completion is complex and multifaceted. The review of the literature that follows on millennials and first generation students, predictive analytics and college retention, and classification of academic disciplines revealed agreement that no single solution can be universally applied to make all students successful in degree completion. The recent change of student support in California, enacted by the Seymour-

Campbell Student Success Act of 2012, has the potential to provide community colleges the opportunity to focus on specific student needs in order to increase student completion.

Social media companies and online retailers use predictive analytics to drive custom content to site users to increase sales, site visits, and support of advertisers (Chen, Chiang, & Storey, 2012). Limited research on using predictive analytics to forecast student outcomes and deliver effective interventions in higher education has been undertaken, in spite of the fact that these institutions are collecting and maintaining an array of data on students and performance. Leveraging existing student information may be effective in developing predictive models to forecast student completion (Smith, Lange, & Huston, 2012). The use of data analytics, also known as "big data," could be used to deliver targeted support services to individual students who need support in completing their programs based on key identifiers in the student's record (Bradley, 2013). This predictive analytic model would enable colleges to use the information contained in the student record to identify those at risk of failing to complete their degree and provide intervention through support services to reduce attrition.

Big data is a term describing extremely large sets of information that may be evaluated computationally to identify patterns, trends, and relationships (Chen Chiang, & Storey, 2012; Church & Dutta 2013; McAfee & Brynjolfsson, 2012; Pavolotsky, 2013). Big data is used in data analytics, the science of evaluating information (data) with the intent of drawing conclusions (Johnson, 2012). Predictive analytics and predictive modeling involve the extracting of information from existing data sets to identify patterns and then using those patterns to predict future outcomes (Smith, Lange, & Huston, 2012).

These same strategies are used by companies such as Amazon, to provide users with suggested products based on their browsing histories (Johnson). The approach for the use of big data is similar in the marketing concept of driving customized content to the consumers on social media sites (Church & Dutta, 2013).

The application of predictive analytics requires that the large volume of available information contained in the student record be researched to determine the predictive relationship between the data points and degree completion. Once the data are categorized the relevant data points, those with predictive relationships, can be integrated into the analytics model. The first year of college has been determined to be critical in the student experience including degree completion (Burns, 2010; Hickman, 2011; Tinto, 2004). The critical nature of the first year experience includes the courses taken during the first term; what is not available is a strategy that allows an institution or a system to categorize courses in a manner that is amenable to big data analytics.

This study provided an opportunity to determine if Biglan's (1973) taxonomy is an effective categorization scheme for first term course failure information in a predictive model. Biglan's seminal work on disciplinary differences was used to classify first year courses as hard/soft, pure/applied, and life/nonlife. The classification afforded review and analysis of the data across types of courses, and ultimately, across institutions. This study's purpose was to address the problem of students not completing degree programs within 6 years by exploring one possible variable in defining potential support needs based on data collected at the community college level with regard to addressing aspects of community college completion rates.

Problem Statement

Limited resources are a persistent and significant issue for community colleges striving to increase student success and degree completion (Adams, 2015; Engle & Tinto, 2008; Hickman, 2011; Summers, 2003). The great economic downturn increased the significance of this issue for community colleges in recent years, but no single solution has been identified (Board of Governors, 2013; Murray, 2010). In order to maximize the impact of available resources to address student completion, it is necessary to identify factors that predict failure to complete in order for schools to take advantage of the data and computing power available to them. Although it is one of higher education's most parsimonious and studied models, the utility of Biglan's (1973) taxonomy to big analytic strategies had not been tested.

Purpose of the Study

The purpose of this non-experimental causal-comparative study was to advance predictive research strategies by determining a tenable categorization scheme that might be included in a predictive analytics model designed to identify the risk of attrition and opportunity to deliver intervening support services to increase student success. Specifically, Biglan's (1973) taxonomy was applied to the diversity of courses students took in their first year during the timeframe for this study. Only students who failed a course in their first term were included in the sample for this study.

In this quantitative study, I tried to determine if a predictive relationship existed between the type of course failed in the first term of study, based on Biglan's (1973) taxonomy, and failure to complete a degree or certificate within 6 years. The independent

variables were the three aspects of type of course failed in the first term: hard/soft, pure/applied, and life/nonlife. The dependent variable was degree attainment (yes or no) within 6 years. Predictive asymmetrical log-linear analysis (LLA) was used to test this question.

Research Questions and Hypotheses

To demonstrate the utility of the Biglan (1973) taxonomy to big data strategies, a single research question was determined: Is there a predictive relationship between type of course failed, as cross-categorized by the dimensions in Biglan's taxonomy, in the first term and failure to complete a degree or certificate within 6 years?

Null hypothesis H_0: There was no predictive relationship between type of course failed in the first term and failure to complete a degree or certificate within 6 years.

Alternative hypothesis H_A: There was a predictive relationship between type of course failed in the first term and failure to complete a degree or certificate within 6 years.

In other words, the true test of the premise of this study was the identification in the post hoc analysis of particular types of failed courses that were overrepresented in noncompleting group.

Theoretical Foundation

This study was grounded in Tinto's theory of retention (1975), which holds that a student's tendency to stay in college is related to the degree to which the student is integrated into the academic and social realms of the college. The current research on

student completion indicated that the issue is complex and multifaceted, and cannot be resolved through a single solution (Burns, 2010; Engle & Tinto, 2008; Hickman, 2011; Metz, 2002; Schroeder, 2013; Summers, 2003; Turner & Thompson, 2014). Identifying the factors related to failure to complete a degree may lead to more rapid, and therefore timely, intervention through the delivery of support services targeted to the specific needs of individual students (Smith, 2012).

Retention is a term used to describe a student persisting in an institution of higher education in pursuit of a degree or certificate (Tinto, 1975). Tinto's theory of student retention identified the relationship between a student's connection with both the social and academic spheres of the college. The greater the student perceives their connection with both the academic and social life of the institution, the more likely the student is to persist to completion of a degree (Tinto, 1975, 1987).

Integration of students into the academic sphere of the institution includes academic preparedness, academic skill support, sufficient levels of academic performance, and individual academic goals (Tinto, 1987). Tinto (1987) determined that individual commitment, or the willingness to expend the effort required to complete a degree, had significant influence on student retention and that individual commitment varies over the course of an academic experience within a college. Further, Tinto determined that the experience of a student within the college was more important to persistence than what a student has experienced prior to arrival at the institution (p.6). A student's academic success, or lack thereof, at the college has more of an influence than prior academic success at the high school level.

Many research findings support the importance of the first year of college in completion of a degree or certificate (Engstrom & Tinto, 2008; Summers, 2003; Turner & Thompson, 2014). It is during this first year that colleges have the greatest opportunity to provide support services to students at risk of failure to complete. This type of information would be especially germane to big data strategies.

Studies have determined that student retention and degree completion can be positively impacted through targeted intervention support (Bean, 1980; Bean & Metzner, 1985; Cabrera, Nora, & Castañeda, 1993; Coppola, 1999; Elkins, Braxton, & James, 1998; Engle & Tinto, 2008; Kalsner, 1991; Maalouf, 1993; Murray, 2010; Spady, 1971; Summers, 2003; Tinto, 1997, 2004, 2008). This study expanded on the research on retention by determining if a predictive relationship exists between the type of course failed in the first term and failure to complete a degree or certificate within 6 years. The intent was that if such a predictive relationship was found to exist, it may be used as a data point in a predictive analytics model designed to deliver targeted intervention support services to students at risk of failing to complete a degree. There was no existing research to support use of Biglan's (1973) taxonomy for the classification of courses in a predictive analytics model. However, the recent emergence of data analytics as a tool for higher education decision making brings opportunity for classification schemes to reduce the number of data points in the model.

Nature of the Study

This quantitative study employed a non-experimental causal-comparative design that allowed me to investigate the extent to which the independent variables (type of first

term course failed) may have impacted the dependent variable (degree completion within 6 years). The independent or construct variables were the types of courses, based on Biglan's (1973) taxonomy hard/soft, pure/applied, and life/non-life, failed in the first term. The dependent or criterion variable, measured to determine differences between types of courses and student demographics, was degree or certificate completion within 6 years. The 6-year period aligned with the student cohort model required for data reporting by California community colleges (Harris, 2015). Categories of types of courses failed in the first term were determined using Biglan's taxonomy.

The study used existing student data from a small, rural California community college district referred to as Valley College District. The de-identified data were provided by the institutional researcher of the college district and included all students who enrolled for the first time during the fall and spring semesters of 2006/07, 2007/08, and 2008/09 and failed a course in their first term. The database identified students who failed a course in the first term and provided all required data for the study including degree and certificate completion and student demographics.

Definitions

The following terms were used in this study:

Completion: Achieving a degree or certificate (CCCCO, 2015).

Failed Course: Student did not earn an A, B, C, or P grade (CCCCO, 2015).

Retention: Student persistence across academic terms and years leading to completion (Tinto, 1975, 1987, 1997, 2004, 2008). In this study, retention refers to student enrollment from one term to the next.

The study focused on students who declared a program of study. Students who were undecided in their major were excluded. The student achieving completion was defined as one who completes the courses fulfilling the requirements of a degree or certificate program within 6 years from the start of their first term of enrollment.

Assumptions

The assumptions in this study represented components that could not be controlled for by me. The following assumptions were made for this study. Courses offered by the participant college could be effectively classified according to Biglan's (1973) taxonomy. Students provided accurate information on their application for admission. The college has kept accurate and complete data during the years 2006 through 2015. Course failure might have indicated an inability to perform academically, which is identifiable in the first year by course failure.

Scope and Delimitations

Smith, Lange, and Huston (2012) proposed that colleges could forecast student outcomes through leveraging existing student information in predictive analytics. In order to construct a predictive model to forecast student outcomes, it is necessary to identify the data elements in the student information system that have a predictive relationship with student outcomes. This study focused on the outcome of degree or certificate completion within 6 years only and was specifically designed to explore the relationship between the type of course failed in the first term and failure to complete a degree or certificate within 6 years. The intent was that Biglan's (1973) taxonomy might be incorporated into a predictive analytics model to identify students at risk of failure to

complete. Biglan's taxonomy provided a structure that reduced the complexity of data in a predictive model to increase the effectiveness in application affording the implementation across institutions.

This study was a secondary analysis of archived data. Use of archived data in an ex post facto study shortened the timeline required and supported replication of the study in other institutions. Additionally, this study did not consider the academic or social integration of students or their level of academic preparedness. The limitations of the study included the small sizes of the participating colleges. The small numbers included in the study may limit the effective generalization of the study and limit the external validity. The colleges were federally identified as a Hispanic Serving Institutions with unduplicated headcount of 10,000 annually.

Significance

This study was designed to determine whether Biglan's (1973) taxonomy can be used in a predictive analytics model to increase degree completion through delivery of targeted student support services. If so, it would support the college effort to direct scarce resources to students with identified risk factors of failure to complete. Previous studies have determined a variety of factors with a positive relationship to failure to complete and determined the effectiveness of a broad range of intervention support programs that increase the likelihood that a student will complete a degree or certificate. If a relationship between type of course failed in the first term and failure to complete a degree or certificate within 6 years existed, then type of course failed in the first term may be incorporated into a predictive analytics model to identify students at risk of

failure to complete. The type of course failed has not been previously used but may provide a broad factor category related to failure complete. This approach will support including new and revised courses into the predictive analytics model as they are developed, based on Biglan's taxonomy. Further, targeted interventions and support services can be delivered to students based on the predictive model outcomes to increase the likelihood of successful degree completion.

The knowledge gained from this study may also increase the understanding of student retention and lead to development of additional intervention support services. The results of this study may change how students at risk of failure to complete are identified and how support services are targeted to the identified student. Community colleges provide access to higher education to historically underserved populations. When provided effective and timely support services, these students achieve success leading to sustained and generational social change in society.

Summary

Decades of research on student retention have identified a variety of factors impacting completion (Bean, 1980; Burns, 2010; Davis, 2010; Hickman, 2011; Spady, 1975; Summers, 2003; Tinto, 1975, 1997, 2008). The complexities of retention and the student information stored by colleges afford the opportunity to apply predictive analytics in addressing the issue of completion. Students at risk of failure to complete a degree or certificate within 6 years may be identified through a predictive model affording the opportunity to drive customized content or intervention support services to increase the likelihood that a student will successfully achieve their educational goal. The purpose of

this study was to investigate the relationship between the type of course failed, based on the Biglan (1973) taxonomy, in the first term and failure to complete a degree or certificate within 6 years to determine the validity of incorporating this strategy into predictive analytics models.

Chapter 1 of this study served to provide an overview of the problem, the background, and goals of the study. The research questions and hypotheses were introduced to build understanding of the purpose and intended outcomes of the study. The theoretical framework was presented as context for the reader to further understand the study and the literature reviewed in Chapter 2. In Chapter 2, an analysis and summary of research and articles relating to retention, classification of academic disciplines, millennials and first generation students, and the science of big data is presented. Chapter 3 presents the detailed methodology of the study, the procedures for gathering and analyzing the existing data, and the instruments used. Chapter 4 discusses data collection and results leading to the interpretation of findings, implications for social change, and recommendations in Chapter 5.

Chapter 2: Literature Review

The open access mission of California community colleges brings with it the challenge of educating students unprepared for the rigors of program completion (Engstrom & Tinto, 2008). Although the California community college system has been successful in efforts to increase the number of first generation students attending college ("First-generation…", 2014), retention persists as a significant issue (Murray, 2010; Schroeder, 2013). This chapter contains a review of research on community college completion, millennial and first generation students, and application of predictive analytics in higher education as a foundation for this study.

In spite of the volume of research on student success and retention, the complexity of student support in higher education makes improving degree completion an ongoing challenge (Cabrera, Nora, & Castañeda, 1993; Engle & Tinto, 2008; Kalsner, 1991; Schroeder, 2013; Tinto, 2004). According to the California Community Colleges Chancellor's Office (CCCCO), California community colleges serve a growing demographic of first generation students who are more likely to receive financial aid, more likely to enroll in developmental courses, and far more likely to be Hispanic than the non-first generation student group (CCCCO, 2014)--all factors associated with attrition. Additionally, the millennial generation, which comprises the majority of California community college enrollments, is characterized by exceedingly close relationships with their parents and an overreliance on communications technology (Elam, Stratton, & Gibson, 2007), which are also associated with attrition. The challenge to practitioners is identifying points in the academic process where intervention is

possible, by leveraging the digital data footprint of these students (Bradley, 2013; Parry, 2012).

California community colleges have engaged in a redesign of support programs and services to increase degree completion (Seymour-Campbell Student Success Act, 2012)). The redesign was focused on institutional data and research-based practices as keys to improving student outcomes (Seymour-Campbell Student Success Act, 2012). Course failure and its relationship to degree completion in community colleges comprise one such potential area to be leveraged and is recognized as a meaningful problem (Engle & Tinto, 2008). In this study of students who failed one course in their first term and then failed to complete within 6 years, the purpose was to identify which types of courses failed during the first term had a predictive relationship with failure to complete a degree within 6 years. Biglan's (1973) taxonomy was used to establish the type of courses failed among the diverse course offerings taken by first term students. If the determination that a predictive relationship exists between the types of courses failed and failure to complete a degree within 6 years, community colleges could support the use of first term course failure as an indicator within a data analytics framework to drive customized support services to students to increase success leading to completion. For example, if a first term failure in a hard/pure/non-life science course (e.g., mathematics) shows a significant relationship with failure to complete a degree in 6 years, then that information could be used to trigger intervention through the custom content delivery of support services for students enrolling in these courses.

For the literature review in this chapter, I included an exploration of research on millennials and first generation community college students, focusing on the challenge of retaining both of these groups of students. Current research on big data and predictive analytics is a second topic explored to demonstrate how these might be used to increase student retention leading to completion. Finally, the review includes research on the role predictive analytics plays in delivering custom content in an educational environment.

The remaining sections of this chapter include the literature search strategy, a discussion of Tinto's (1975) theory on retention and the various aspects of research that support it as the theoretical foundation for the variables tested in this study, and a literature review that includes millennial college students, first generation college students, and predictive analytics as a tool for increasing student success and completion. The chapter concludes with a summary of the findings of the research and description of the gap in the literature that this study seeks to address.

Literature Search Strategy

The literature and research reviewed was comprised of peer-reviewed articles identified through entering key search terms into the following databases: EBSCO Host, Education Research Complete, and Proquest Dissertations and Thesis. The peer-reviewed articles, journals, and dissertations were directly related to degree completion of community college students. The preponderance of journal articles was obtained from ERIC. Key search terms used to develop the literature review included: *retention, completion, community college, success, data analytics, big data, predictive analytics,* and *academic discipline*. The key terms entered included a variety of Boolean parameters

available. Searches that combined the term *community college* with *retention* or *success*, for example, served to identify research most closely related to this study. The literature review extended from 1973 through 2014; however, the early research was limited to Tinto's (1975) seminal work on student retention and Biglan's (1973) taxonomy of college courses.

Theoretical Framework

The earliest theoretical models of attrition of community college students were attributed to Spady, Tinto, and Bean and Metzner (Summers, 2003). These early models of student attrition focused on identifying the numerous factors that relate to a student's decision to either remain in college or to drop out. The theoretical framework discussion that follows provides an overview of work that led to the seminal work of Tinto (1975) and his theory of retention which serves as the theoretical framework for this study.

Retention Theorists and Models

Developing actionable knowledge that could be applied to student retention calls for a focus on outcomes and the ability to analyze and apply insights (Johnson, 2012). Extending this actionable knowledge to the challenge of community college degree completion may assist California community colleges in increasing success leading to completion. Murray (2010) stated, "We need to know the students that we are serving and how to serve them well" (p. 35). Student attrition models suggest that completion can be increased by strengthening the engagement of students with the institution and providing learning support services (Bean, 1979; Spady, 1971; Tinto, 1987).

Durkeim's (1951) research on suicide framed the development of the earliest models of student attrition (Bean, 1981). The key element of Durkheim's theory reflected in the student attrition models is that of connection. When a person shared values with a group, was connected, the individual was less likely to commit suicide (Durkeim). Similarly, the student attrition models of Spady (1971) and Tinto (1987) identified integration and community membership, with the college, as influences on reducing attrition. These early models served as the foundation for contemporary student attrition theories and research (Elkins, Braxton, & James, 1998; Metz, 2002).

Spady (1971) developed the first theoretical model of the dropout process. Analogous to Durkeim's (1951) theory that shared group values and the support of friends reduce the risk of suicide, Spady's model highlighted social and academic integration as key influences in the decision to drop out of college. Ambition is the key concept of what Spady termed the aspiration-attainment process, the process leading to degree completion. Spady's theory considered educational aspiration and college attainments as outcomes of a complex interrelated social dynamic of school and family. Spady included grade point average (GPA), family socioeconomic status, intelligence, parent's aspirations, extracurricular activities, achievement motivation, and achievement values as variables in the study. Of these variables, academic achievement was found to be the strongest influence in the ambition process (Spady).

Spady (1971) framed his research on the concept that support and friendship decreased the likelihood that a student would drop out. Spady noted that dropout decisions occurred as a longitudinal process influenced by the student's background

characteristics. In addition, Spady added the educational variables of grade performance and intellectual development as influences in the dropout decision-making process because each of these was believed to increase social integration. Spady proposed that social integration and interaction would increase satisfaction, thereby leading to decreased attrition. Spady's model was reflected in the student departure model presented by Tinto (1975).

Building from Spady's (1971) model of the dropout process, Tinto (1987) proposed another model of student attrition that included social and intellectual communities as an essential element of effective retention. Tinto's model of retention included the influence of institutional commitment to students, educational commitment by students, and institutional commitment and educational choice in addition to the social and intellectual communities. Tinto's model will be the primary framework for this study.

Tinto (1975) developed the most widely cited and tested model of the student attrition process (Metz, 2002, p. 3; Summers, 2003, p. 3). Tinto based his model on the parallel processes that were found between students who dropped out and individuals who committed suicide. Tinto applied Durkheim's (1951) theory on suicide, which holds that suicidal tendencies increased in people who were not socially integrated, to student attrition (Summers, 2003).

Tinto's (1975) model includes the background characteristics of family, individual attributes, and the experience of pre-college schooling and elements that interact with one another. Further, these elements were expected to influence the students' commitment to their educational goals and the institutional commitment to the

students. Commitment in academics leads to higher grade achievement and intellectual development (Elkins, Braxton, & James, 1998; Tinto, 1987). Higher grade achievement and intellectual development lead to increased academic integration which continues the cycle and leads to even greater student commitment (Tinto, 1987). In this manner, student educational commitment and/or educational goals reduce the likelihood that a student will dropout. In Tinto's model, institutional commitment is expected to provide additional interaction between student groups and faculty, which fostered greater social integration, which in turn, was expected to produce increased institutional commitment. Tinto's model of student attrition has a circular design on the student commitment side and along with the institutional commitment sphere works together toward reducing the result of dropping out.

Since his seminal work on student attrition, Tinto (1975) continued to explore retention, persistence, and completion from a variety of perspectives. Later, Tinto (1987) introduced the "paradox of institutional commitment" (p. 11), the paradox being that institutions willing to encourage students to leave were also those that were more likely to have students stay (p. 11). Tinto explained that the paradox of institutional commitment can be argued to hold the key to effective action for institutions seeking to increase student retention. Tinto's research suggested that institutions who were more committed to the education of students would be willing to tell students when it was in their best interest to leave and that those were also the institutions that were likely to have students more committed to them and therefore, more likely to persist until degree completion.

The paradox of institutional commitment seems to not fit the mission of an open access environment that undergirds the contemporary community college system (Tinto, 2008). These colleges are expected to provide educational access and opportunity to all who desire to attend, which precludes telling a student when it is in their interest to leave. Further, the colleges are expected to address the challenge of student completion in a manner that reflects the values of open access and equity. Although Tinto's (2008) paradox does not appear to operate in the California community college system because of open access mission, his model of attrition remains the appropriate foundation for this research study because it has been used effectively for other similar studies on student retention including the influence of late registration on retention at a community college (Maalouf, 2012).

Extensions of Tinto's Model. Tinto (1987) proposed the need for institution level assessment of student departure. Tinto noted that the complexity of student departure was further complicated by the specific institutional context where it transpired. Metz (2002) also identified the importance of institution level exploration of student departure, which strengthens the fit of Tinto's work with the purpose of this study. Bean (1981) synthesized the student attrition models of Spady (1971) and Tinto (1975) to develop a causal model of student attrition. The resulting synthesized model of student attrition identified four classes of variables: background variables, organizational variables, environmental variables, and attitudinal and outcome variables (Tinto, 1987). Bean suggested that each of the variables have an influence on intent to leave, a precursor to dropping out.

The classes of variables included in Bean's (1981) model were congruent with the variables in Tinto's (1975) model of student attrition. However, the Bean model was intended to inform the institution of the types of challenges the university could expect when admitting students with specifically identified characteristics, whereas the Tinto (1987) model informed colleges of the effect of their institutional commitments on attrition and the support needed for students to successfully remain enrolled and persist toward completion.

Bean (1987) expanded attrition research by applying a model of the employee turnover in work organizations to institutions of higher education. This expansion to higher education ultimately provided the foundation for Bean's (1981) position that Durkheim's (1951) theory of suicide was not a solid theoretical basis for attrition models. Where the attrition models of Spady (1971) and Tinto (1975) relied on social connection as the key element of the intent to leave, Bean's (1987) made no such assumption. Bean sought to test the power of the employee turnover in work organizations model as a model for student attrition and to rank order the variables according to their influence on attrition. The study suggested that male and female students left the institution for different reasons (Bean). Although institutional commitment was determined to be important to both males and females, the variables influencing the commitment varied (Bean). Also, the study determined that the second most influential variable for men and women was academic abilities; for men, this variable was their university GPA and for women it was their academic performance at the high school level (Bean).

Additionally, Bean's (1987) study of the causal attrition model indicated that institutions that provide opportunity for freshmen students to develop personal, intellectual, creative, and inter-personal skills had lower levels of attrition. And finally, the perceived quality of the education the student was receiving was determined to be one of the most important variables for both males and females in influencing their institutional commitment, which led to degree completion (Bean). As a result, Bean's suggested that the main task of future studies was in the identification of missing determinants in the causal attrition model (p. 32).

In the context of this study, community colleges have an unprecedented opportunity to increase completion rates through understanding the factors that influence student success and the institutional factors that influence student retention, persistence, and goal attainment (Burns, 2010). Determining if a relationship between type of first term course failure and completion exists will increase the breadth of factors found to influence student success. This then increases the opportunity to deliver appropriate support services to increase the likelihood of degree completion in accordance with Tinto's model for enhancing retention and graduation (2004).

Tinto's Theory in Current Research. Tinto's (1975) seminal work on attrition theory has been widely used as the framework for research on student retention and completion. Further, Tinto continued to engage in large scale research studies, such as the Pell Institute for the Study of Opportunity in Higher Education (Engle & Tinto, 2008). This national study of low-income, first generation students, revealed a number of challenges that made it more difficult for this population of students to be successful in

college. Low-income, first generation students were disproportionately from ethnic and racial minority backgrounds and were less likely to be academically prepared for college level work. These students were found to be older, less likely to have financial support from their family, and to more often have obligations outside of college, including children and jobs, that limit their opportunity to engage in extracurricular activities and the full college experience. Engle and Tinto's attrition model research revealed that these factors lead to the decision to drop out.

Kalsner (1991) examined the following four themes of student attrition identified with Tinto's (1987) path analysis model: (a) uncertainty about what to expect from college and its rewards, (b) transition and adjustment problems, (c) financial difficulties, and (d) lack of academic preparation. Kalsner's study results suggested that the student selection process was one means of increasing persistence; however, it was noted that altering the selection process to admit students more likely to persist might create a more homogeneous student population rather than increasing the completion rates of all students. This study became part of the growing body of research that validated Tinto's (1987) theory of student attrition in a variety of institutions with differing student populations.

Cabrera, Nora, and Castañeda (1993) conducted a study to determine the extent to which Bean's (1981) student attrition model and Tinto's (1975) student integration model could be combined to increase understanding of the decision process leading to attrition. Variables in the study included three indicators of family and friends support; financial support, academic integrations, social integration, institutional commitment, and

institutional fit (Cabrera, Nora, & Cateñeda, p. 130). The results revealed that the role of environmental factors in the decision to drop out was far more complex and extensive than proposed by Tinto (1975). The findings supported Bean's (1981) position that environmental factors should be taken into consideration when designing a model of student retention. The results of Cabrera, Nora, and Castañeda's (1993) study further indicated that merging the two theories could improve understanding of students' persistence.

Summers (2003) explored the relationships between community college student enrollment and registration behaviors and student characteristics and academic outcomes. The study included Tinto's (1975) student attrition model, which identified student characteristics related to specific enrollment and registration behaviors, including late registration. Summers examined interrelationships among academic outcomes and sought to determine if enrollment and registration behaviors were predictive of student academic outcomes and the findings suggested that there were significant differences between the characteristics of students who persisted as compared to students who dropped out. Students who persisted were of traditional college age (under age 25), White, pursuing transfer, and not eligible for financial aid assistance (Summers, p. 9). Summers concluded that further study of community college registration and enrollment patterns had the potential to provide valuable understanding and identification of students likely to depart.

Metz (1999) challenged the premises of Tinto's (1975) student attrition model, most notably the exclusion of community college students in the study's population. Building on the work of other researchers, Metz (1999) explored the fit of Tinto's model

with 2-year college students. Among the research studies cited by Metz was the study on college persistence by Cabrera, Nora, and Casteñeda (1993). Metz's (1999) findings confirmed that Tinto's (1975) model was reliable when applied to 2-year college populations similar to when applied to traditional student populations. These current research studies support the use Tinto's theory as the framework for this study.

Rationale for Choice of Theory

Tinto's (1975) student attrition model has been applied in numerous research studies (Burns, 2010; Metz, 2002; Summers, 2003) making it the most widely tested and accepted model. The model has provided insight into the role of integration of students into the social and academic realms of the college (Tinto, 1975, 1987, 1997), enrollment and registration behaviors (Burns, 2010), and persistence of first-time, full-time college freshmen (Elkins, Braxton, & James, 1998). The model has been determined to be effective when applied to 2-year college students as well as 4-year university students in determining the factors of student attrition (Metz, 2002).

Tinto (1987) determined that approximately three fourths of all dropouts leave during the first year of college. The first term and first year of college have been important areas of research as practitioners strive to improve community college degree completion. The community college student population is increasingly comprised of first generation college attenders and millennials; two populations with identified characteristics associated with attrition (Adams, 2015; Elam, Stratton, & Gibson, 2007; Engstrom & Tinto, 2008; Tinto, 2004, 2008; Turner & Thompson, 2014; Varallo, 2008).

First generation students in the California community college system are more likely to be female, older, and more reliant on financial aid than non-first generation students ("First-generation…", 2014). Further, they are more likely to be less academically prepared thereby placing into developmental classes (Davis, 2010) and generally complete fewer units of credit ("First-generation…"). First generation students are less likely to successfully complete courses and less likely to persist into year 2 (Davis, 2010).

As shown in this review, although the retention is a relatively simple variable to consider in a study (students are retained or not), the factors that may predict this discrete outcome are quite diverse and often interactive. Importantly, what is different *now*, than when Durkheim and others started looking at the challenge of retention, is the potential impact of technology on understanding the complexity in these predictors, particularly with millennial students who are first generation. These days, researchers are able to track nearly every interaction that a student has with and institution and pair it with all the information they have on a student using predictive analytics. The literature review begins with research on millennial first generation students, followed by research on the use of predictive analytics and, finally, Biglan's (1973) taxonomy, which will be used to classify first year courses in addressing the research question of this study: Does failure in a particular type of course in the first term predict failure to complete in 6 years?

Literature Review

The literature review that follows begins with an exploration of research on characteristics and retention of millennials and first generation community college

students, for population focus of this study. This is followed by current research on big data and predictive analytics to demonstrate how they might be used to increase student retention leading to completion. The role of predictive analytics in delivering custom content in a higher education environment is explained. Finally, Biglan's (1973) taxonomy as the source for categorizing the potential predictors of community college completion is presented.

Millennials and First Generation College Students

Probably no single institution in education has had a stronger impact on the quality of people's lives than the community college system. The open access mission of community colleges makes them an ideal institution for students with a broad spectrum of educational experience and preparedness. Community colleges are serving growing populations of students characterized as millennials and first generation attenders (Becker, 2009; Elam, Stratton, & Gibson, 2007; Engstrom & Tinto, 2008; "First-generation…", 2014; Jones & Healing, 2010; Tinto, 2008; Turner & Thompson, 2014; Varallo, 2008). Importantly, although some community college students are both millennials and first generation attenders, this population is only one subset of students. There are students that are millennials and non-first generation attenders. There are first generation students that are not millennials. Definitions of millennials and first generation students follow in the narrative review of research on these identified populations.

Characteristics of both millennials and first generation students have been identified in an effort to increase understanding of current students as a means of addressing the challenge of increasing community college completion (Murray, 2010).

Millennials and first generation students are the focus of numerous recent research studies; however, this literature review is focused on studies designed to identify characteristics and factors of these specific populations related to retention, completion, and attrition.

Millennials were defined by the Pew Research Center (2014) as the 60 million individuals born after 1980, a generation for which no chronological endpoint has been set. Elam, Stratton, and Gibson (2007) analyzed research on millennial students in the context of college admissions, advising, and student support. The authors synthesized the primary research and identified the characteristics of millennials beginning with the overarching characteristic of having been raised in an era of real time media coverage (p. 21). Millennials grew up watching the war in Iraq unfold with live, continuous footage and commentary: they experienced the shootings at Columbine High School and the resulting school lock-downs and increased campus security; and they watched the initiation of impeachment proceedings against a sitting President of the United States (Elam, Stratton, & Gibson, 2007, p. 21). Despite this new world of immediate and live coverage of world events, millennials were raised by parents who provided steady support and protective concern about their safety, education, and extracurricular activities (p.21). These factors may have increased the ability of millennials to structure their time, follow rules, and multitask (Elam, Stratton, & Gibson, p. 22). However, millennial students presented challenges to educators because of their exceedingly close relationships with their parents who often assume participatory roles in their education (p. 22). During this time, the term *helicopter parent* was coined.

Millennials were also raised with technology and social networks (Elam, Stratton, & Gibson, 2007). They are the first generation of students with experience and expectation of communication at all times, and they have engaged in group projects and worked together on team projects throughout their primary and secondary education experiences (Elam, Stratton, & Gibson, 2007, p. 24). Millennial students have experienced curricula that encouraged rote learning leaving many without the skills necessary to be critical thinkers and demonstrate self-reflection (p.22). Even though the technical skills of millennial students enable them to find the answers, they may lack the ability to demonstrate mastery of concepts through application of knowledge.

Becker (2009) and Jones and Healing (2010) examined millennial behavior to frame a generational context for understanding agency, choice, and technology use. In the classroom, millennials desired to be engaged, participative, and have active control of their education which Becker noted was a good fit with a learner centered classroom environment. While millennials were the first generation to grow up with computers and were referred to as digital natives; they exhibit a range of technology prowess (Becker, 2009: Jones & Healing, 2010). Further, the information literacy of millennials was found to range from quite limited to knowledgeable and experienced in identifying reliable sources. According to Becker, while students have the ability to multitask with diverse types of digital media, they may not all have the technology and information skills needed for educational success (2009, p. 352).

Jones and Healing (2010) determined that the universities had the necessary conditions for millennials including access to computers and broadband networks.

Student interviews revealed that students engaged with the technologies they believed they were required to use in their courses even using the technologies more than they were required to (Jones & Healing, 2010, p. 351). A theme common to most of the interviews was the distracting nature of social networking sites, even when they appeared to be focused on an educational task. Similar to Elam, Stratton, and Gibson (2007), Jones and Healing observed the technology savvy and connected characteristics of millennials. As noted by Elam, Stratton, and Gibson (2007), it is important to recognize that not all millennials have been afforded equal levels of financial, personal, and social support. Scott-Clayton (2011) pointed out that the evidence that a problem existed was very strong, but the evidence of what interventions work best for whom and under what circumstances was limited. Colleges continue to be challenged to create and shape environments, processes, and programs that nurture and sustain students across the college experience to improve student success, retention, and completion rates (Schroeder, 2013). The research on millennials suggested that these students are characterized by familiarity with ubiquitous technology and expectations of connectivity (Elam, Stratton, & Gibson, 2007; Jones & Healing, 2010). Millennials were found to desire an engaged and participatory learning experience and to exhibit a range of technology prowess (Becker, 2009; Gibson, 2007).

First generation students as a research category includes both millennials and older students. As a group, first generation students are defined as those whose parents had no college experience (First-Generation students in the California Community College System, 2014). As a group, they have been characterized by multiple factors

related to lower success and completion rates ("First- Generation…", 2014). Further, Scott-Clayton (2011) determined that the evidence of completion as a persistent issue is very strong, but the evidence of what interventions work best for whom and in what circumstances remains limited.

The data on California community college first generation students was aligned with the findings of Tinto (2008). Tinto (2008) explored the increased access to higher education for low income students and discovered that while more low income students were entering college, fewer were successfully completing their program of study and ultimately earning a four-year degree. Consistent with most definitions, first generation college students have been defined by the California Community College Chancellor's Office as students whose parents have no college experience ("First-Generation…", 2014).

Since 2012, California community colleges have been required to collect, via the common community college application, and report parent education data. A research brief was released in September 2014 presenting a state system overview of first generation students. Highlights from the summary included findings that first generation students were slightly more likely to be female and of similar of age in comparison to non-first generation students. First generation students were far more likely to be Hispanic than non-first generation students. First generation students were more likely to receive financial aid and enroll in developmental courses than non-first generation students. First generation students were slightly less likely to successfully complete courses and equally likely to persist from fall-to-fall as non-first generation students. The

profile of the first generation student in California community colleges includes the above enumerated factors that have been determined to have a negative influence on completion. In particular, lower socioeconomic status has been found to have a strong relationship with attrition (Engle & Tinto, 2008).

Degree completion among community college students is subject to a wide array of variables that influence student persistence. The variables warrant exploration for consideration for use in a predictive analytics model. The research question that guided this study is: Does failure in a particular type of course in the first term predict failure to complete in 6 years?

Predictive Analytics and College Retention

Predictive analytics, big data, business intelligence, and data analytics are terms used to describe practices of using large volumes of data to predict outcomes (Chen, Chiang, & Storey, 2012; Church & Dutta, 2013; Johnson, 2012; McAfee & Brynjolfsson, 2012; Pavolotsky, 2013) The science of predictive analytics can assist community colleges in identifying students who are unlikely to complete a degree and provide a new opportunity for delivery of support services targeted to address specific indicators of attrition. The unresolved issue of student retention in community colleges can be addressed through application of the student retention model proposed by Tinto (1975) to identify the factors that predict a student at risk of failure to complete. Once identified, the factors of attrition can be tracked by student information systems at community colleges. Identification of the factors of failure complete creates the predictive analytics framework necessary to affect the use of big data at the student level in alignment with

Summers' (2003) finding that college computer systems can be utilized to increase student persistence. The big data research findings of Summers (2003) and Church and Dutta (2013) support the application of Tinto's (1975) model of student attrition for the purpose of researching the relationship between type of course failed in the first term and completion of a degree or certificate within 6 years.

The potential of big data can be described as the extent that institutions can capture, understand, and take effective action from the information elements within the data (Church & Dutta, 2013, p. 23). Predictive analytics has become an approach that leverages massive amounts of information through the integration of the data into actionable outcomes. The application of predictive analytics in higher education is emerging and not yet facilitating identification of students at risk of failure to complete.

Chen, Chiang, and Storey (2012) found that multivariate statistical analysis, such as regression, factor analysis, clustering, and discriminant analysis, have been used in a variety of business applications. Alternatively, optimization techniques, derived from the management science community, have been applied to enhance web crawling. Based on the interest in, and success of, the data mining and statistical analysis in business intelligence, advanced processes and technologies have been developed to support text analytics, web analytics, network analytics, and mobile analytics.

McAfee and Brynjolfsson (2012) studied the value, in terms of business performance, of being data driven in decision-making. They found that that data driven decision-making was not limited to the published accounts, but instead transforming the business environment similar to the findings of Church and Dutta (2013). The premise of

this study is that these tactics can be used for colleges and universities. The problem is developing a way to make sense of the data.

Church and Dutta (2013) found it was critical for leaders to manage the change to data driven decision-making effectively to gain the full benefit of the use of big data. Leaders must change the institutional understanding of the role of leadership and company culture around decision-making. Further, they predicted that every sphere of business activity will be impacted by this movement toward data driven decision-making. Predictive analytics has been described as the science behind the custom content delivered to Amazon customers and Facebook users (Parry, 2012). Smith, Lange, and Huston (2012) proposed that by leveraging existing student information, colleges could build statistical models, or predictive analytics, to forecast student outcomes. The concept of predictive analytics and big data was noted by Bradley (2013) as an opportunity to track students on a case-by-case basis and design individualized interventions (p. 7).

Smith, Lange, and Huston (2012) presented a case study on a community college using learning analytics and the development of predictive models to identify at-risk students based on dozens of identified key variables. The study by Smith, Lange, and Huston centered on Rio Salado College in Tempe, Arizona, a college that serves over 60,000 students annually. Over half of the students enrolled at Rio Salado College took courses via online delivery. The study was designed to identify the factors that demonstrated a significant statistical correlation with final course outcomes. The research study included every online course offered at the college; however, the results presented were limited to one freshman level, full online course. The sample consisted of 539

students enrolled in the course between May and December of 2009. The study demonstrated strong correlation between online course activity markers and course outcome, where online course activity markers included; login to the course homepage, viewed course syllabus, opened an assessment, completed an assessment, viewed grade book comments from instructor, viewed assessment feedback from instructor, opened a lesson, requested due date change, and selected a custom calendar option. Higher numbers of activity markers correlated to higher course grade. The results of Smith, Lange, and Huston showed that frequency of course login, site engagement, pace, and assignment grades could serve as effective predictors of course success. These predictors were determined to be effective as early as the 8th day of class providing a large window of opportunity for delivery of support interventions if needed.

Predictive analytics are used broadly by online retailers and social networking platforms to drive customized content to the end user (Parry, 2012). There are limited studies on the application of predictive analytics in higher education, yet colleges maintain volumes of individual student data that may be used to identify key factors of risk of failure to complete and ultimately provide opportunity to deliver customized intervention support services to increase success and completion.

What may be missing in utilizing this "fire hose" of data pouring into institutions' databases is a way to organize and make sense of it for practitioners tasked with responding to any insights gleaned from it. Specifically, data elements that indicate risk of failure-to-complete must be identified and integrated into these predictive models. Biglan's (1973) taxonomy will be used to classify courses in this study in an effort to

determine if type of course failed in the first term has a predictive relationship with failure to complete a degree within 6 years.

Biglan's Taxonomy

Biglan (1973) developed a framework for classifying academic disciplines to examine the relationships between subject matter characteristics and department organization within universities. Biglan employed multidimensional scaling of subject matter characteristics at a large, urban university and at a small college located across the country and compared the results from these unlike institutions. A total of 186 faculty members of the large university and the entire faculty of the small college, approximately 70, participated in making judgments about the similarities of academic areas (Biglan, 1973, p. 196). The faculty participants were selected to include as diverse as sample as possible across the 36 areas of interest with no less than three and no more than five judges in any area. The judgments were made on 18 academic areas that were common to both institutions.

Biglan (1973) plotted the scaled results for each institution, which provided a visual representation of the faculty judgment of the characteristics of subject matter of academic areas. From the graphic representation and the scaled results, Biglan (1973) concluded that three dimensions appeared to characterize academic subject matter. The dimensions include: the degree to which a paradigm existed, hard/soft; the degree of concern with application, applied/pure; and concern with life systems, life/nonlife (Biglan, 1973, p. 202). Biglan determined that these characteristics had a direct effect on the organizational structure and work of an academic department.

Table 1

Sample of Academic Domains Classified According to Biglan's Two Main Dimensions

	Hard	Soft
Pure	Mathematics Chemistry Astronomy	Sociology Art English
Applied	Computer Science Civil Engineering Horticulture	Accounting Communication Business

Biglan's taxonomy, or classification model, has been used most often since 1996 for classification in researching differences among academic disciplines (Jones, 2011). Colleges and universities were regularly organized around academic disciplines based upon rationale that linked the research and instruction of the academic department (Del Favero, 2014). Influence in the academic profession also stemmed from disciplinary foundations. Lack of hierarchical structure of the teaching profession led to a structure in higher education based on influence and persuasion (Del Favero, 2014, p. 1). Disciplinary affiliations afforded development of incentives and forms of cooperation around subject matter and expertise. Since the 1850s, and still common today, colleges and universities are organized around this type of disciplinary grouping (Jones, 2011).

Pike and Killian (2001) conducted a study on the differences in students' college experiences and learning outcomes using Biglan's (1973) taxonomy to identify academic domains. The study used a conceptual design, including two distinct aspects of student integration, involvement and integration, to determine if the students' perceptions of the

college environment were directly related to gains in student learning and intellectual development (Pike & Killian, 2001, p. 433). The study's participants included 2,000 seniors at a Midwestern university. The students were designated by major as either a pure or applied discipline according to Biglan's classification (Pike & Killian, 2001, p. 435). Participants completed a survey that measured their involvement in academic activities and the degree to which they had made gains in learning (p. 438). The results of the research study were ambiguous (p. 446). Students in applied disciplines were determined to have more positive views of the college learning environment than did students in pure disciplines. According to Pike and Killian, the type of academic discipline, as determined by Biglan's taxonomy, was the strongest predictor of student perception. Where Pike and Killian's study results were congruent with characteristics noted by Biglan, a more recent study resulted in stronger correlation of academic disciplines and epistemological beliefs.

Biglan's (1973) taxonomy was used by Schommer-Aikins, Duell, and Barker (2003) in their research examining student beliefs about the nature of knowledge, or epistemological beliefs, across academic domains. The study of 152 university students employed three questionnaires that measured epistemological beliefs about mathematics, social sciences, and business (Schommer-Aikins, Duell, & Barker, 2003, p. 347). Two analyses were undertaken to determine the correlation of epistemological beliefs between mathematics and social science and between mathematics and business. Schommer-Aikins, Duell, and Barker's regression analysis revealed that social science and business epistemological beliefs were significantly related to the corresponding mathematical

epistemological beliefs (p. 358). Based on the analysis, Schommer-AIkins, Duell, and Barker proposed that both social science and business epistemological beliefs can be relied upon to predict mathematical epistemological beliefs. The results of this study indicated that epistemological beliefs of undergraduate students were moderately domain general, which indicated that there was a limited degree of alignment of epistemological beliefs when results were analyzed using Biglan's classification (p. 360). However, contrary to Biglan's taxonomy of academic domains which predicted stronger correlations between mathematics and social sciences, the study by Schommer-Aikins, Duell, and Barker revealed stronger correlations between mathematics and business (p. 360).

The two later studies by Pike and Killian (2001) and Schommer-Aikins, Duell, and Barker (2003) supported the use of Biglan's (1973) taxonomy for classification of courses in research. Pike and Killian's study results include the finding that the relationships among all of the constructs in the conceptual model were stable across academic environments allowing the differences between pure and applied disciplines to be directly assessed (Pike & Killian, 2001, p. 446). The results of the study by Schommer-Aikins, Duell, and Barker similarly supported the use of Biglan's classification of academic domains as a viable construct for evaluation in research. I did not find a current study of type of first term courses failed as a predictor of degree completion using Biglan's taxonomy during the literature review process; however, the research reviewed supports the use of Biglan's taxonomy for academic classification in

determining the relationship between type of course failed and failure to complete a degree or certificate within six years.

Summary and Conclusions

Completion is a persistent challenge for community colleges (Murray 2010; Schroeder, 2013). Even as access has been increased, student success leading to completion of degrees and certificates remains a recognized challenge for colleges (Engle & Tinto, 2008; "First-Generation…", 2014). Tinto's (1975) retention theory served as a framework for many studies (Burns, 2010; Metz, 2002; Summers, 2003) that increased understanding of the factors influencing student completion. Millennials and first generation students pose an additional challenge for community colleges as they have been characterized by expectations of social challenges and other factors that have been determined to influence failure to complete (Adams, 2015; "First-Generation…", 2014; Murray, 2010; Schroeder, 2013: Varallo, 2008). As Hickman (2011) noted, institutions can only hope to positively affect student persistence through understanding what influences it.

The literature search and review did not identify a current study of type of first term courses failed as a predictor of degree completion using Biglan's (1973) taxonomy during the literature review process; however, the research reviewed supported the use of Biglan's taxonomy for academic classification in determining correlation between type of course failed and failure to complete a degree or certificate within 6 years. An understanding of the relationship between the type of course failed in the first term and failure to complete a degree or certificate within 6 years would assist in determining the

validity of using type of course failed in the first term as a data value in the predictive analytics model resulting in an increased understanding of the factors of attrition. This study was constructed to determine the validity of using course type, according to Biglan's taxonomy, as a data point in a predictive analytics model that would afford delivery of customized support services to students deemed likely to fail to complete.

Chapter 3 describes the quantitative research design that was used for this study. The quantitative data for this study included student demographic data that is archived in the community college student information system. It also includes enrollment information, course grade data, and degree and certificate attainment information.

Chapter 3: Research Method

The specific purpose of this study was to determine if the type of course failed in the first term has a predictive relationship with failure to complete a degree or certificate within 6 years. A determination of this kind of predictive relationship is important in development and implementation of a predictive analytics model that will allow colleges to increase retention through delivery of student support services based on factors indicating a student is at risk of attrition. Use of a predictive analytics model will allow colleges to direct limited resources to students at the highest risk for failure to complete. In this chapter, I provide background and rationale supporting the choices made related to the methodologies for this research study. In the chapter, I also include an overview of the research design, threats to validity, use of archived data, and data analysis. The archive data and procedures for data procurement are discussed and I conclude the chapter with an overview of the issues associated with the methodology of this study.

Research Design and Rationale

A nonexperimental quantitative design was employed for this study. The quantitative approach is appropriate when seeking to determine if a relationship exists between the dependent and independent variables (Cresswell, 2003), specifically for this study, the type of course failed in the first term and failure to complete a degree within 6 years. The quantitative research method supports evaluation of relationships through the analysis of data including outcomes. In this study, the outcome or dependent variable was degree completion. A student either completed or did not complete a degree within 6 years. The independent variables were the type of course failed in the first term, as

described by the three dimensions of the Biglan's (1973) taxonomy. The independent variables were also categorical in nature. The research question for this study required a quantitative design because neither a qualitative nor a mixed method approach may be used to determine if a relationship exists between these variables (Cresswell, 2003). Only a quantitative design supports drawing scientific conclusions through hypothesis testing.

Research Question and Hypotheses

This non-experimental design study was a secondary analysis of student data to address the research question: Is there a predictive relationship between the type of course failed, as cross-categorized by the dimensions in Biglan's (1973) taxonomy, in the first term and failure to complete a degree or certificate within 6 years? A null hypothesis and an alternative hypothesis stemmed from this research question:

Null Hypothesis (H_0): There is no predictive relationship between the type of course failed in the first term and failure to complete a degree or certificate within 6years.

Alternative Hypothesis (H_A): There is a predictive relationship between the type of course failed in the first term and failure to complete a degree or certificate within 6 years. Specifically, a post hoc analysis to identify the particular types of failed courses that are overrepresented in noncompleting group will be conducted.

Methodology

Within the quantitative research methods, I identified an ex post facto, causal comparative nonexperimental design for this study. Causal comparative studies are used to determine the effect of one variable on another (Campbell & Stanley, 1963). The causal comparative design is nonexperimental and includes an independent variable that

is not manipulated by the researcher (Creswell, 2003). An experimental design would have required random sampling which did not fit the research question; additionally, the ex post facto design fit with the use of archived data. Correlational design is similar to causal comparative; however, it requires that the independent and dependent variables to be continuous. All variables in this study are categorical and not continuous, eliminating correlational design as an option. The causal comparative design was the appropriate choice for this study because it afforded me the opportunity to explore the relationship between the type of course failed in the first term with degree completion within 6 years to answer the proposed research question.

Population

California community colleges enroll over 2,000,000 students annually (Board of Governors, 2013). The 113 colleges in the state system face the challenge of increasing degree completion as set forth in the Seymour-Campbell Student Success Act of 2012 (Board of Governors, 2013). First term students are of particular interest because of the research suggesting the relationship between first year experience and degree completion as noted in Chapter 1. California was selected for this study because of the number of students served by the 72 community college districts (CCCCO, 2015). California community colleges are the largest higher education system in the United States of America (CCCCO, 2015).

The theoretical population is the population to which researchers desire to generalize the results of study (Trochim, 2006). For this study, the theoretical population was all first term students enrolled in a community college in the United States.

According to Trochim, the accessible population is that to which the researcher has access. In this research study, the accessible population was first term students enrolled in California community college in the academic years 2006/07, 2007/08, and 2008/09. The colleges selected as representative are referred to as the Valley College District in this study. Valley College District was selected based on the demographics which reflect demographics of colleges across the California community college system. The selected years provide a timeframe that enables the first term students a 6 year window for completion.

Sampling and Sampling Procedures

This study used a nonrandom, convenience sample comprised of the census archived data from the selected institution. A census study includes all members of the accessible population that fit the sampling frame (Johnson & Christensen, 2004). Although random sampling is prized for the high degree of internal integrity (Tipton, 2014), it was not reasonable for this study. A census sample eliminates the step of generalizing from the sample of study participants to the accessible population as required for sample technique studies, and therefore, increases the external validity of the study results (Johnson & Christensen, 2004).

Not all members of the accessible population were included in this study; however, the population was selected to reflect the scope of this study. In particular, the sample was limited to students who declared a program of study and failed a course in the first term. Students who failed a course that does not fit the classification of courses according to Biglan's (1973) taxonomy were also excluded. The exact steps followed to

extract these data are discussed in the next section. Students who failed more than one course in the first term were excluded from the study; this served to maintain the integrity of the relationship between the specific type of course failed and failure to compete.

It is critical that a sampling strategy provide sufficient data to answer the research question (McMillan & Schumacher, 2001). The sampling method needs to be congruent with the type of research, the research hypotheses, fiscal resources available, the number of variables studied, the data collection method, and the degree of accuracy desired (McMillan & Schumacher, 2001, p. 177). This study was designed to allow for generalizing from the study sample to the larger population. For log-linear analysis, an expected sample size of 10 per cell is a minimum requirement for analytical power (Salter, 2003). This study included an analysis of a 2x2x2x2 contingency table, which necessitated a minimum of 160 respondents or students in the sample. A preliminary screening of the data set indicated that the census sampling strategy exceeded the minimum requirement, which was roughly 400.

Archival Data

This study used archival student data from Valley College District, which has a current student population of 8,900. Valley College District agreed to provide archival new student data from 2006/07, 2007/08, and 2008/09 subsequent to IRB approval by Walden University (#09-04-15-0016453). An institutional researcher at Valley College District accessed and compiled the data from Datatel Colleague (a product of Ellucian) to ensure that student identity was masked in the records (i.e., did not include their name, address, social security number, or any other identifiers). The institutional researcher

agreed to maintain the identification key for a period of 5 years. A computer generated identification number was used to further ensure that student identity was protected and that the data could be sorted and tracked accurately during analysis. The information was provided to me in a spreadsheet.

The archival data spanned 3 academic years from 2006/07 through 2008/09. The data fields included the variables selected for analysis: course failed in the first term of enrollment and if a degree or certificate was awarded within 6 years. The data did not include the student name or other identifier to ensure confidentiality. Only an informal local process was required for access to the Valley College District student data; the institution released the data upon receipt of the Walden University IRB number. I initiated the process via email and phone call. The archival data was be retrieved from the student information system in which all required documentation of student records for the institution are retained. All data was provided electronically via password protected systems. The data remained electronically secured throughout the analysis process and will be maintained according to ethical research practices.

Operationalization of Variables

All the variables in this study were nominal-level and dichotomous. The dependent or outcome variable, completion, was scored as *yes* or *no.* The independent or predictor variables were operationalized by the three dichotomous categories of *hard* or *soft*, *pure* or *applied*, and *life or non-life* from the Biglan (1973) taxonomy. Courses that did not have attributes that met the criteria for categorization using Biglan's taxonomy were excluded from the study. Excluded courses included physical education activity

courses, such as walking for fitness and weightlifting; courses designed to support student success including study skills and career planning; and skill building courses, such as card dealing that are specific to casino employment.

Data Analysis Plan

To demonstrate the utility of the Biglan taxonomy to big data strategies, a single research question was determined: Is there a predictive relationship between type of course failed, as cross-categorized by the dimensions in Biglan's (1973) taxonomy, in the first term and failure to complete a degree or certificate within 6 years? Log-linear analysis was selected for analysis of the predictive relationships in the data. An extended version of the chi-square test of independence, LLA is an effective technique for nominal-level variable research studies (Salter, 2003) and was selected over other evaluation tools including linear and logistic regression. In this study, the use of predictive LLA (P-LLA) supported drawing inferential conclusions regarding the predictive relationship of a dependent variable, based on the categorical status of the independent variables. It allowed me to determine if the type of course failed (hard/soft, pure/applied, or life/non-life) had a predictive relationship with degree completion within 6 years. SPSS statistical software was used for analyses.

Several approaches were implemented to reduce the chance of making a Type I error (rejecting the null hypothesis in error). The first approach was the analysis of all variables in a single analysis to avoid overanalyzing the same data. Further, a census sample was used as a method of maintaining the largest sample size reasonable for the study, thereby assuring that the requisite sample size for P-LLA for determining

predictive relationships was met. This approach to contingency table analysis required a minimum cell size of 10 to support power (Salter, 2003) and the estimated sample was 1,500. Finally, a Bonferroni technique (U.S. Department of Commerce, 2013) was applied for post analysis to counteract the problem of multiple comparisons and further reduce risk of Type I error. The Bonferroni method fit well with this study because it is valid even in cases of unequal sample sizes. Because a targeted intervention required additional resources and the sample size is viewed as robust, a conservative threshold criterion ($\alpha = .01$) was selected to be used to determine statistical significance of the test results.

The archive data provided by Valley College District was coded by the researcher of the district, including the classification of courses according to Biglan's (1973) taxonomy. The students in the study sample were divided into two groups representing the category of their dependent or outcome variable of degree or certificate completion (CC) within 6 years of their first term of attendance. Those scores were cross-tabulated across the dichotomous predictor variables of course type. The following representative notations were used for the variables; *hard* or *soft* (HS), *pure* or *applied,* (PA), and *life* or *non-life* (LN). This cross-tabulation was represented by a 2x2x2x2 contingency table. Before moving to the main analysis, the table was screened for anomalies. It was expected that a few anomalies would be identified. The table was also screened for cells with zero observations, although one or two cells with no observations are not generally a problem. Given the size of the sample, this outcome was not expected.

The log-linear method relies on the likelihood ratio chi-square (L^2) instead of the more familiar Pearsonian X^2. The benefit of the L^2 statistic is the opportunity to assign and assess the significance of the amount of chi-square attribute associated with each of model components similar to methods of analysis based on the general linear model (e.g., ANOVA) The LLA approach provides additional benefit in that model components are all additive supporting easy determination of a parsimonious explanatory solution. For this study, I applied the asymmetrical LLA technique of P-LLA, which incorporated both dependent and independent variables. The following equation represents the null LLA model and serves as the starting point for this study.

HS * PA * LN = CC

Following the P-LLA analysis and assuming the null LLA model was significant, analysis of standardized residuals (SR) was used to determine if any cells in the contingency table were overrepresented. SRs were distributed as *z*-scores, which made them especially amenable to interpretation. Because the central aspect of this study concerned students who were not retained after 6 years, only the configurations of Biglan types in that group's 2x2x2 table were examined in the post hoc analysis. A Bonferroni technique (U.S. Department of Commerce, 2013) was used to ensure elimination of over interpretation of the multiple comparisons in the post hoc analyses. For the post hoc analyses, the a priori alpha (.01) divided by 8, the number of the cells in the 2x2x2 contingency table for non-completers, presented the criterion ($p < .00125$) for interpretation of SRs.

Threats to Validity

Any research design should address factors that may result in unexpected outcomes in the study results. These factors are known as threats to validity. Threats to validity may be internal, those that affect the sample or study within the variables, or external, those that prevent the generalization of results (Johnson & Christensen, 2004). A research design should include measures to reduce the threats of internal and external factors. A research design should also include measures to reduce the threats of measurement validity and conclusion validity. For this study, the threat of measurement validity was minimized by the exclusion of any tool other than the coding by the researcher. The internal, external, and conclusion threats to the validity of this study are explored below.

Threats to External Validity

Threats to external validity are those aspects that prevent the generalization of the study results to the broader population (Johnson & Christensen, 2004). Of the six threats to external validity associated with causal comparative research design (Cresswell, 2003; Martella, Nelson, Morgan, & Marchand-Martella, 2013), two were applicable to this study. First, the generalization across participants was addressed through acknowledgement of the limitations of a census as the sample in the results section of the study. Valley College District was selected because it is representative of other community colleges in California both in demographics of students served and in the types of first year courses offered. The characteristics of the institution selected for the census, Valley College District, allowed for comparisons to the general population to be

made, although the results may not be generalized to the broader population with a degree of precision associated with scientific research.

This study used a census sample to increase the external validity of the research results. A census sample supports generalizing from the sample to general population (Johnson & Christensen, 2004) and maximizes external validity. Additionally, four threats to external validity were addressed by the design of this study. The first, verification of the dependent variable, was a precise determination of degree or certificate completed within 6 years. This dichotomous measure minimized threat to external validity. The second threat, posttest sensitization, was not applicable to this study because there was no pretest and posttest in the design. The third threat, interaction of time of measurement and treatment effects, was similarly minimized through the study's design, which was a secondary analysis of archival data. The fourth threat, interaction of history and treatment, required that I not generalize the study results to past scenarios, future scenarios, or other individuals. I took care to adhere to this guideline.

Threats to Internal Validity

Internal validity is the degree of confidence to which the results of a study can be attributed to the variables in the study and not other influences (Campbell & Stanley, 1963). This study did not employ an experimental design, rather a causal comparative study and therefore was not subject to each of the typical internal threats to validity associated with experimental design. Specifically, causal comparative design is susceptible to the selection threat to internal validity. This limitation to the study was addressed through use of census.

The ex post facto nature of the study eliminated influence of the research project on results as a threat to internal validity. Because the student completion outcome had already been determined, there was no opportunity for the study of archival data to influence the results. Secondary data analysis may be subject to limitations including archival data that are outdated, unreliable, incomplete, or from an institution that is not similar to the general population (Royse, Thyer, Padgett, & Logan, 2006). In this study the threats to internal validity from use of archival data were limited through my familiarity with and comprehensive knowledge of the institution, the courses, and student information system. Additionally, I had direct communication with and access to the institutional researcher and executive leadership of Valley College District for consultation. To further minimize internal threats to validity in this study the statistical tests of significance were appropriately selected for the design and scope of the study and the study was be grounded in theory.

An additional threat to internal validity is the operationalization of the independent variable (Cresswell, 2003). This aspect was minimized through the use of Biglan's (1973) taxonomy for the dichotomous classification of variables; however, not all courses may easily be classified based on the defined attributes. The threat was further minimized by excluding courses that could not be classified using Biglan's taxonomy. The studies reviewed in Chapter 2 provided extensive guidance on appropriate coding of these courses, as well.

Threats to Conclusion Validity

I used SPSS for statistical analyses to ensure accuracy and reduce threats to conclusion validity based on computational error. Further, I maintained awareness of researcher expectancies and approached drawing conclusions with caution to minimize threat of construct validity. Finally, the goal of any data analysis strategy is to use techniques that align with the nature of the variables in the study while respecting the overall research design; a predictive study using nominal level variables. The P-LLA was selected to assure the integrity of the conclusions I drew because the technique is both parsimonious and powerful in this context.

Ethical Procedures

This study adhered to the highest level of professional practice in educational research. The institution and points-of-contact remained confidential and were not included in the results of the study. If the college requests the results of this study, they will be provided. The identities of the individual students included in the archival data will remain confidential. My contact information, including phone numbers and email address, were been provided to the institutional researcher at the participating college to facilitate communication and provide opportunity to request the results of this study.

This study required Institutional Review Board (IRB) approval at Walden University #09-04-15-0016453. The IRB approval ensured that the proposed research study would treat all human subjects ethically. The approval was attached to the request for access to archival data at the institution included in the study. The institution provided

a letter of consent to participate. A copy of all records will be stored and secured electronically for a minimum of 5 years.

Summary

Chapter 3 provided an outline of the research design and methodology planned for this study, including the research question, hypotheses, variables, plan for analysis, threats to validity, and ethical procedures. Chapter 4 provides the data collection, data analysis, and discussion of findings from the study. Chapter 5 will conclude the study with an interpretation of the findings, limitations of the study, recommendations for further study, implications for social change, and a conclusion.

Chapter 4: Results

The purpose of this study was to advance predictive research strategies by determining a tenable categorization scheme that might be included in a predictive analytics model designed to identify a risk of attrition and an opportunity to deliver intervention support services to increase student success. The approach was a nonexperimental causal comparative study designed to determine if a predictive relationship exists between the type of course failed in the first term of study, as categorized by the Biglan's (1973) taxonomy, with failure to complete a degree or certificate within 6 years. The independent variables were the three aspects of the type of course failed in the first term: hard/soft, pure/applied, and life/nonlife. The dependent variable was whether or not there was degree or certificate attainment within 6 years. This chapter provides a description of the data collection, the secondary analyses of those data, and a summary of the results addressing the research question.

Data Collection

This study employed a nonrandom census sample of archived student records from Valley College District whichwas selected as representative of others in California. The quantitative data were compiled by the institutional researcher of the participating community college district in response to my request. I received the electronic data file from the researcher on September 9, 2015. The data from two small, rural community colleges were included representing first time students who failed a course in the first term of enrollment. The data were from the six primary academic terms, fall and spring, in academic years 2006/07, 2007/08, and 2008/09. The colleges selected were determined

to serve students representative of California community college populations. Demographic information was also included in the dataset including gender, age, and race/ethnicity of the students.

As described in Chapter 3, a few key choices were made to arrive at the data set that was analyzed in this study. Students who failed more than one course in the first term were not included. Of the original sample of 1,979 unduplicated students, 220 exclusions were made to ensure that all courses met the criteria for categorization using Biglan's (1973) taxonomy. These courses included physical education activity courses, such as basketball and walking for fitness, and guidance classes emphasizing career exploration and developing study skills. A total of 1,759 students were included in the study after all exclusions were made to the data set. There were no missing values.

The census sample of 1,759 students exhibited the demographic characteristics illustrated in Table 2, which appeared largely consistent with the range of community college cohorts in California (California Community Colleges Chancellor's Office, 2015). The participants ranged in age from 11 to 71 years of age at the start of the first term of attendance. The mean age of the sample was 22.65 years of age. The extremes of the youngest and oldest student participants were consistent with vast span of ages among the nontraditional populations served by community colleges.

Table 2

Demographic Characteristics of Study Sample (N=1,759)

Demographic Variables	*n*
Age	
Under 20	930
20-24	390
25-39	325
40 and over	112
Unknown	2
Gender	
Female	927
Male	813
Unknown/Decline to state	19
Ethnicity	
African-American	184
American Indian/Alaskan Native	19
Asian	61
Filipino	27
Hispanic	921
Pacific Islander	15
Two or more races	1
Caucasian Non-Hispanic	397
Unknown/Decline to State	134

The study sample was not proportional to the sampling frame with respect to gender and ethnicity, although the differences were small and not viewed as a threat to external validity of the results. The respondents were 52.7% female and 46.2% male, while the participant college population was 59.5% and 39.1% respectively (California Community Colleges Chancellor's Office, 2015). With regard to ethnicity, the study sample was reflective of individuals in the sampling frame. In both the sample and the college student populations, Hispanic students comprised the largest ethnic group.

Caucasian Non-Hispanic was the second largest ethnic group in the sampling frame and in the study sample. The African–American group was proportionally larger in the study sample than in the college population. The number of American Indians/Alaskan Natives, Asians, Filipinos, and Pacific Islanders was proportional across the study sample and the participating college population.

Of the 1,759 students who failed a course in the first term of enrollment, 111 (6.3%) completed a degree or certificate within 6 years and 1,648 (93.7%) did not complete a degree or certificate within 6 years. The observed low rate of completion does not align with the trends identified in prior research on completion at the community college. Nationally, first time student community college completion rates have been around 30% at the end of 3 years (Burns, 2010). According to Bailey and Morest, (as cited in Burns, 2010) at the 8 year mark, 39% of first time community college students have earned a degree or certificate (p. 34). Engle and Tinto (2008) reported that 60% of low-income, first generation students who leave postsecondary education before completion do so at the end of the first year.

The course-level data were coded according to Biglan's (1973) taxonomy, resulting in each course failed in the first term being categorized by the three dichotomous predictor variables of course type, which were cross-checked by the methodologist on my supervisory committee. The courses include hard (23%) or soft (77%), pure (54%) or applied (46%), and life (31%) or non-life (69%). Because no baseline data exists on Biglan types in community college courses, I am unable to know if this distribution was representative.

When cross-tabulating degree outcomes with the courses by type, one of 16 cells of the 2x2x2x2 contingency table had zero observations. Based on the overall small number of observations in the completer group, this outcome was not unexpected. The cell representing *hard/applied/life* had zero instances of students who completed a degree within 6 years after failing a course in the first term. To support analyses of the data, I chose to use a delta of .50 for that cell in the data analysis. Because this tactic complicates estimation of various log-linear models, I also examined the L^2s for the marginal associations (i.e., not including one of the three independent variables) to provide another view of the effects where a delta was not needed.

Results

This study addressed one research question: Is there a predictive relationship between the type of course failed, as cross-categorized by the dimensions in Biglan's (1973) taxonomy, in the first term and failure to complete a degree or certificate within 6 years? The four variables were treated at the nominal level, and LLA was used to find the most parsimonious model that fit the observed frequencies in the 2x2x2x2 table (reported in Table 3). As this LLA was asymmetrical, the analysis began with evaluation of the null model (L^2 (7, N = 1,759) = 16.809, p = .019). The associated probability was not as pronounced as hoped, but did suggest that perhaps a simpler model might emerge (discussed next and in Chapter 5). Because an examination of SR was proposed however, that information is also included in Table 3. No SR in the noncompleting group approach the stipulated level of significance ($p < .00125$), which was likely attributable to the fact that they comprised 94% of the sample.

Table 3

Observed and Expected Frequencies, and Standardized Residuals for Completion by Biglan Type (N=1,759) from the Null Model

Hard/Soft	Pure/Applied	Life/Nonlife	Observed	Expected	SR	*p*
Degree						
Hard	Pure	Life	6	6.310	-0.124	.45
		Nonlife	20	12.368	2.170	.02
	Applied	Life	0	1.451	-1.205	.11
		Nonlife	2	5.111	-1.376	.08
Soft	Pure	Life	3	5.427	-1.042	.15
		Nonlife	37	35.969	0.172	.43
	Applied	Life	28	21.266	1.460	.07
		Nonlife	15	23.096	-1.685	.05
No Degree						
Hard	Pure	Life	94	93.690	0.032	.49
		Nonlife	176	183.632	-0.563	.29
	Applied	Life	23	21.549	0.313	.38
		Nonlife	79	75.889	0.357	.36
Soft	Pure	Life	83	80.573	0.270	.39
		Nonlife	533	534.031	-0.045	.33
	Applied	Life	309	315.734	-0.379	.35
		Nonlife	351	342.904	0.437	.33

Note. SR = standardized residual, *p* = probability.

The lack of error terms in LLA often allows a researcher to determine a more parsimonious explanatory model however. Toward that end, each component L^2 was examined to understand its contribution to understanding the observed frequencies in the full contingency table (see Table 4). The three independent variables were found not to be significant as main effects. Instead, most of the residual chi-square (14.496) was to be found in the interactions of the independent variables. Specifically, the most parsimonious explanation of differences in whether a student completed would seem to be best predicted by the interaction of *life/nonlife* and *pure/applied* aspects of the courses that were failed ($p = .009$). Examination of the SRs of this more parsimonious explanatory model did not yield any over or underrepresented cells.

Table 4

Summary Table for the Asymmetrical Log-Linear Analysis of Degree Completion by Course Type

Model (Source)	Residual L^2	Component L^2	Marginal L^2	*p*
Null		16.809		.019
HS	16.345	0.407		.523
PA	15.381	1.370		.242
LN	16.488	0.288		.592
HS*PA	9.157	5.346	5.903	.015
HS*LN	10.787	3.724	2.568	.109
LN*PA	6.937	7.553	6.871	.009
HS*PA*LN		1.683		.195

Note. p = probability. Probabilities reported for the marginal L^2s.

This smaller model may be a more parsimonious explanation for the observed data. The most likely confounding factors to not obtaining the desired results were the high number (93.7%) of students in this sample who failed a course in the first term and failed to complete a degree or certificate within 6 years, and the low numbers of particular types of courses taken at the undergraduate level as described in the Biglan model. In particular, there were four courses coded as *hard/applied/life* and 38 courses coded as *soft/applied/non-life.* The difference in numbers of courses in each category contributed to the variance of observed results in Table 3. The category *hard/applied/life* included four courses and 23 observations. The category *soft/applied/nonlife* included 38 different courses and 366 observations. While there is no recognized base line for expected frequencies of Biglan types in a sample of community college courses, the study sample was not evenly distributed across the eight types.

Summary

Based on the analyses conducted, the original null hypothesis could not be fully rejected for the research question: There was no predictive relationship between the type of course failed in the first term and failure to complete a degree or certificate within 6 years, in the full model. When the *hard/soft* variable was removed however, significant effects were noted in the analysis of the interactive model of the independent variables of *life/nonlife* and *pure/applied.*

In Chapter 4, I described the data collection, study sample, and results of my study. In Chapter 5 I provide a summary of the study with an interpretation of the findings. I also present implications for social change and recommendations for further

study on first term course failure and the relationship with failure to complete a degree or certificate within 6 years.

Chapter 5: Discussion, Conclusions, and Recommendations

This chapter provides a brief review of the purpose, design, process, and results of the study to frame the discussion of the findings. The findings will be interpreted in relationship to the research question of the study, the theoretical framework, and the recent body of research that served as the foundation for the study. The implications of the study for social change will be presented and the chapter will conclude with recommendations for action and further study of community college completion.

Approximately 45% of all students in higher education are enrolled in community colleges (Giancola & Davidson, 2015). Community college students are more likely to be individuals of color, the first generation of their family to attend college, and have lower incomes when compared to students enrolling in other types of colleges (Engle & Tinto, 2008). Engle and Tinto (2008) found that low-income first generation students were nearly four times more likely to leave higher education after 1 year than their peers without these risk factors.

Decades of educational research have focused on identifying a variety of risk factors of attrition. These risk factors include: limited academic and social integration (Bean, 1980; Bean & Metzner, 1985; Burns, 2010; Cabrera, Nora, & Castañeda, 1993; Tinto, 1975, 1987, 1997, 2004, 2008; Spady, 1971), inadequate student support services (Engstrom & Tinto, 2008; Murray, 2010; Tinto, 2008), academic preparedness and performance (Elkins, Braxton, & James, 1998; Kalsner, 1991; Summers, 2003) first generation students (Oldfied, 2007), and the millennial generation (Becker, 2009; Elam, Stratton, & Gibson, 2007; Jones & Healing, 2010; Scott-Clayton, 2011; Varallo, 2008).

These studies' findings have fostered design of student support services and interventions to increase student completion; however, the challenge of degree completion within 6 years, a national benchmark, persists.

This study was developed to explore a strategy for supporting application of predictive analytics to forecast student completion. In a manner similar to that of social media companies and online retailers, community colleges maintain vast amounts of data on each student. The diversity of the data within the student information systems presents a challenge for higher education institutions. Classification schemes provide a means of reducing the data diversity and paving the way for implementation of predictive analytics models. This study provided an opportunity to determine if Biglan's (1973) taxonomy could serve as an effective categorization scheme for first term course failure information in a predictive model. A categorization scheme would provide opportunity to reduce course data into a smaller framework of the course type, reducing the number of data points in an analytics model. Identification of students at risk of failure to complete would provide an opportunity for colleges to provide targeted support services to increase student success and retention.

One research question was proposed in this study: Is there a predictive relationship between the type of course failed, as cross-categorized by the dimensions in Biglan's (1973) taxonomy, in the first term and failure to complete a degree or certificate within 6 years? The accessible population for this study included first term students enrolled in a California community college in the academic years 2006/07, 2007/08, and 2008/09. Valley College District was selected as representative based on the

demographics and types of first term courses offered by the institution which reflect those of colleges across the California community college system. The selected years provided the timeframe required for first term students to have a 6-year window for completion.

The data set was screened according to the criteria outlined in Chapter 3 to ensure validity and generalizability of results. The final study sample (N = 1,759) was coded according to Biglan's (1973) taxonomy producing a 2x2x2x2 contingency table. LLA was applied and the null model for the asymmetrical design was fitted and tested for statistical significance. Further exploration was warranted, even though the null hypothesis was not rejected, to determine a simpler model to explain the observed data. In the end, the most parsimonious model was the one for interaction between *life/nonlife* and *pure/applied.* In the next section, the results will be interpreted within the context of the observed data and previous research on community college completion.

Interpretation of the Findings

This study produced three key findings that necessitate further discussion. First, as a group, the independent variables did not have a significant result in predicting completion within 6 years as originally expected. Instead, a simpler model involving the interaction of two independent variables did show a significant predictive relationship to completion, although no cells were statistically over or underrepresented. Third, the vast percentage (93.7%) of students who failed a course in the first term, also failed to complete a degree or certificate within 6 years, which confounded the LLA.

Tinto's (1975) student attrition model has served as the theoretical foundation for numerous research studies (Burns, 2010; Metz, 2002; Summers, 2003). The model has

been determined to be effective when applied to community college students and university students in identifying the factors of student attrition (Metz, 2002). This study used the Tinto model not as a strategy for identifying factors of attrition, but instead as a framework for determining the efficacy of applying Biglan's (1973) taxonomy in a predictive analytics model of student risk of failure to complete. Community colleges use student information systems to maintain student records. These information systems contain a breadth of data points that could be incorporated into a predictive analytics model. Biglan's taxonomy was selected as a categorization scheme for course type to reduce the volume of types associated with courses.

Biglan's (1973) taxonomy has been used most often as the categorization schema for research studies to determine differences among academic disciplines since 1996 (Jones, 2011). Pike and Killian (2001) found that students in applied disciplines had more positive views of the college learning environment than did students in pure disciplines. Similarly, Jones (2011) determined that the organization of colleges and universities around the type of discipline grouping afforded by Biglan's (1973) taxonomy supports cooperation around subject matter and expertise of faculty. The results of these studies included significant differences in results along the independent variables of Biglan's taxonomy. However, in this study, the L^2 associated with the null model for the three categories as a group, did not reach a level of acceptability. Possible reasons for this finding are discussed below.

Because LLA allows the researcher to identify and remove extraneous variables that are not contributing to understanding the observed data, a more parsimonious model

did emerge when I examined all of the components more closely. Specifically, an interaction of the dimensions of *pure/applied* and *life/nonlife* (PA * LN) were determined to have a predictive relationship with degree completion (L^2 (3, N = 1,759) = 6.871, p = .009). As was the case with the null model, no standardized residuals were statistically significant. Therefore, based on this analysis, I could not say that a particular combination of these two aspects of the Biglan model was the variable that was associated with failure.

The study results suggested that the use of Biglan's (1973) taxonomy as a classification schema for courses failed in the first term was not a completely effective predictor of failure to complete a degree or certificate within 6 years. The consideration of a classification schema was intended to reduce the large number of data in the course field of a predictive analytics model; this study suggests that further research with a more balanced sample is needed to determine if Biglan's (1973) taxonomy is an effective classification scheme for use in a predictive analytics model. Other data points will be needed in these predictive models.

As noted in Chapter 4, 96.7% of the students in this sample who failed a course in the first term, failed to complete a degree or certificate within 6 years. This outcome was unexpected and produced a dataset that was not balanced. The lack of balance in the dataset likely made detecting the influence of the independent variables difficult. Further, the group that failed a course in the first term and completed a degree within 6 years was so small that it may be a better predictor than the added nuances offered by Biglan. The application of Biglan's (1973) taxonomy as a schema for categorizing courses failed in

the first term as a predictor of failure to complete a degree within 6 years requires further testing. The study observations suggest that there could be a simpler variable in a predictive model, but it was not confirmed by these results.

In summary, as a group, the three independent variables did not significantly predict completion within 6 years, likely due to distributional differences in the sample. A more parsimonious model did emerge, but again, the distribution did not allow for the identification of specific types of problematic courses. Finally, the majority (93.7%) of students who failed a course in the first term also failed to complete a degree or certificate within 6 years, which suggests that simply failing a course is a better metric for predictive analysis. The findings of this study provide implications for positively impacting social change through increased understanding leading to improved intervention support services for students to both prevent failure in the first term and for those who fail a course in the first term of community college enrollment.

Implications for Social Change

Community colleges provide access to higher education that changes lives and changes generations. Expanding understanding of community college completion and predictors of failure to complete will impact the largest segment of higher education serving the historically underrepresented most at-risk students. This study provided insights into the relationship between first term course failure and failure to complete a degree within 6 years contributing to social change in the practice of student support services leading to increased community college degree completion. The findings of this study challenge institutions of higher education, individual students, and American

community colleges to reconsider the first term of enrollment as it holds social change implications. As the institutions that serve the largest population of higher education students in the United States, community colleges must assess the effectiveness of practice to ensure the support services provided to students lead to success and completion. Community colleges serve historically underserved populations and growing numbers of first generation college students, demographics noted to be at risk of failure to complete; it is imperative that institutions provide the strongest support systems possible fostering individual self-worth through achievement of success and completion.

National reports have indicated that although community colleges have successfully addressed the goal of access to higher education for all who desire to attend, only 11% of low-income, first generation students earn a bachelor's degree compared to 55% of their more advantaged peers attaining bachelor's degrees (Engle & Tinto, 2008). In California, only 59% of the first time students who began in fall 2013 persisted into fall 2014 ("First-Generation…", 2014). Student completion is a complex and multifaceted challenge and understanding the factors leading to attrition will facilitate improvement of any institution's capacity and preparedness for increasing completion leading to social change.

This study was designed to support development of a predictive analytics model that would afford colleges the opportunity to identify students at risk of failure to complete and deliver targeted intervention support services. The predictive analytics model would mirror the tools used by social media companies and online retailers to deliver custom content to site users (Chen, Chiang, & Storey, 2012). Although the

findings of this study did not unequivocally support use of Biglan's (1973) taxonomy in the predictive analytics model, the results reinforced the evidence that the act of failing of a course in the first term, regardless of its type, may have the strongest predictive relationship with failure to complete (Burns, 2010; Hickman, 2011; Tinto, 2004) -a simpler variable in predictive modelling.

Of the 1,759 students who failed a course in the first term of enrollment, only 111 (6.3%) completed a degree or certificate within 6 years. The observed low rate of completion did not align with the trends identified in prior research on completion at the community college; however, this study was limited to those who failed a course in the first term. National data indicates community college completion rates are around 30% at the end of 3 years (Burns, 2010). However, these data included all students, both those who failed a course in the first term and those that do not fail a course.

The finding of a possible relationship between course failure in the first term and failure to complete a degree or certificate within 6 years holds opportunity and implications for the practices of community colleges. Practices must shift from having support services available in general, to targeted intervention in support of student success to reduce course failure in the first term. Further, it provided the opportunity to see the importance of increased intervention and support when a student does fail a course in the first term. As noted by Trainor (2015), community colleges' long history of innovation makes them well-poised to address many of higher education's problems today. This includes improving completion.

Limitations of the Study

The sample in this study was not random; instead, a nonrandom census sample of two small, rural community colleges forming a small college district was used. The college district selected for the study was selected to be representative of California community colleges in the demographics of students served and the types of first term courses offered. The total sample population of students who declared a program of study and failed a course in the first term was 1,979. Individuals were excluded if they failed a course that did not fit the classification of courses according to Biglan's (1973) taxonomy.

The sample size for the study after exclusions was 1,759, which fit the requirements of LLA. One limitation was in the huge difference in size between two groups in the independent variable. While an asymmetrical log-linear analysis controls for this fact statistically, the undersampling of completing group made it difficult to draw conclusions about the noncompleting group, the one of interest in this case.

The sample size was large enough to be statistically successful, but the zero cell in the 2x2x2x2 observed table required a delta of .05 to conduct the LLA. This finding was likely due to the fact that a community college did not have as diverse a range of courses from different disciplines as might be observed across all of higher education. A larger and more diverse sample would decrease the likelihood of a zero observation cell and ensure greater validity and generalization of results, but might not align with institutional level predictive analysis.

Recommendations

The results of this study did not completely support the use of Biglan's (1973) taxonomy as a classification schema in a predictive analytics model for student completion; however, the results of this study did provide opportunity for further research in two distinctly different aspects of community college completion. The first is related to the original purpose of this study--the identification of factors that predict failure to complete as part of the design of a predictive analytics model. The second is further research on the topic of first term course failure in the community college.

Although Biglan's (1973) taxonomy did not prove to be valuable as a classification schema in this research, a larger scale study could produce a different result. This study could be replicated across a larger population allowing for desegregation of results along the characteristics of gender, age, income level, and ethnicity. Additional studies using P-LLA could build on this study resulting in a stronger conclusion that Biglan's taxonomy is not a useful in this application or disproving the findings of this study thereby supporting use of Biglan's taxonomy in a predictive analytics model. Fortunately, lessons from the private sector in using predictive analysis may help guide these efforts in higher education.

Similarly designed research studies could be used to test other data points for use in a predictive analytics model for completion at the community college. These studies could launch a new trend in educational research that specifically examines the predictive relationship between variables, or data points, in the student information system and degree completion to build and strengthen a predictive analytics model. It is beneficial

that the research in the private sector on use of predictive analytics can assist efforts to implement predictive analytics in higher education.

The most pressing area for future research, based on the results of this study, is the impact of first term course failure on college completion. The national and state focus on student success and completion has increased the resources available to provide support services that increase completion. Further study of the relationship between first term course failure and failure to complete a degree or certificate within 6 years will assist in understanding the practices and policies in community colleges that have room for improvement in the drive for increased student success and completion.

Conclusion

Community colleges serve the largest number of students pursuing higher education the United States today. These students are historically underprepared, economically disadvantaged, and more likely to be first generation students than those who attend 4-year colleges and universities. Community colleges are challenged to increase success and completion of the students most at risk for failure to complete. By understanding the interactions between student characteristics and completion, a predictive analytics model that will support the targeted delivery of intervention support services to students most at-risk of failure to complete can be developed. The use of predictive analytics will support efficacy in the delivery of support services that lead to student success and completion. Based on the results of this study, direct attention to the impact of course failure in the first term would be a substantial first step.

References

Adams, C. (2015). Community college students need more support to be successful, research finds. *Education Week.* Retrieved from http://www.edweek.org/ew/contributors/caralee.adams_3652335.html

American Council on Education, InsideTrack, National Association of Student Personnel Administrators, National Student Clearinghouse, & University Professional and Continuing Education Association. (2015). *National study of non-first-time students shows full-time enrollment may not be appropriate for all.* Retrieved from http://www.insidetrack.com/2015/01/20/national-study-of-non-first-time-students-shows-full-time-enrollment-may-not-be-appropriate-for-all

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12,*155-187.

Bean, J. P, & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4), 485-508, 520-530.

Becker, C. H. (2009). Student values and research: Are millennials really changing the future of reference and research? *Journal of Library Administration, 49*(4), 341-364. doi:10.1080/01930820902832454

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57*(3), 195-203.

Board of Governors of the California Community Colleges. (2013). *Revisions to Title 5 regulations: Student success and support programs.* Retrieved from

http://extranet.cccco.edu/Portals/1/ExecutiveOffice/Consultation/2013_agendas/April/regs_student_success.pdf

Borrero, N. (2011). Shared success: Voices of first-generation college-bound Latino/as. *Multicultural Education, 18*(4), 24-30. Retrieved from ERIC.

Bradley, P. (2013). Numbers crunchers: Promise, pitfalls of big data dominate innovations conference. *Community College Week, 25*(15), 6-7.

Burns, K. (2010). At issue: Community college student success variables: A review of the literature. *Community College Enterprise, 16*(2), 33-61.

Cabrera, A., Nora, A., & Castañeda, M. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *Journal of Higher Education, 64*(2), 123-139.

California Community Colleges Chancellor's Office (2015). Data mart. Retrieved from datamart.cccco.edu

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton Mifflin Company.

Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165-1188.

Church, A. H., & Dutta, S. (2013). The promise of big data for OD: Old wine in new bottles or the next generation of data-driven methods for change? *OD Practitioner, 45*(4), 23-31.

Coppola, W. E. (1999). *The relationship of community college student demographic and pre-enrollment background variables with persistence and retention.* (Doctoral dissertation). Retrieved from http:digital .library.unt.edu

Cresswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed method approaches.* Thousand Oaks, CA: Sage Publications, Inc.

Davis, J. (2010). *The first-generation student experience: Implications for campus practice, and strategies for improving persistence and success*. Sterling, VA: Stylus.

Del Favero, M. (2014). Academic disciplines–Disciplines and the structure of higher education, discipline classification systems, discipline differences. Retrieved January 8, 2015, from http://education.stateuniversity.com/pages/1723/Academic-Disciplines.html

Digital community colleges and the coming of the "millennials": Report of major findings from the Center for Digital Education's 2004 Digital Community Colleges Survey. (2004). *Technical Horizons in Education Journal*, *32*(3), 14.

Duell, O. K., & Barker, S. (2003). Epistemological beliefs across domains using Biglan's classification of academic disciplines. *Research in Higher Education*, *44*(3), 347-366.

Durkheim, E. (1951). Suicide: A study in sociology (J.A. Spaulding & G. Simpson, Trans.). Glencoe, IL: Free Press. (Original work published 1897).

Elam, C., Stratton, T., & Gibson, D. (2007). Welcoming a new generation to college: The millennial students. *Journal of College Admission*, 195, 20-25.

Elkins, S. A., Braxton, J. M., & James, G. W. (1998). *Tinto's separation stage and its influence on first-semester college student persistence.* Paper presented at Annual Forum for Institutional Research, Minneapolis, MN. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED424799

Engle, J., & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. *Pell Institute for the Study of Opportunity in Higher Education.* Retrieved from http://files.eric.ed.gov/fulltext/ED504448.pdf

Engstrom, C., & Tinto, V. (2008). Access without support is not opportunity. *Change, 40*(1), 46-50.

First-generation students in the California Community College System. (2014). Retrieved from http://extranet.cccco.edu/Portals/1/TRIS/Research/Analysis/First-Generation%20Students%20in%20the%20California%20Community%20College%20System.pdf

Giancola, J., Davidson, E., & Jack Kent Cooke Foundation, (2015). *Breaking down walls: Increasing access to four-year colleges for high-achieving community college students.* Retrieved from http://cdn2.hubspot.net/hub/281753/file-2369123443-pdf/JKCF_Break_Down_Walls_WP.pdf?t=1421942512492&t=1422564691076

Harris, B. (2015). Goals for the California Community College System (system goals) background and methodology. *California Community Colleges Chancellor's Office.* Retrieved from:

http://extranet.cccco.edu/Portals/1/TRIS/Research/Accountability/System%20Goals%20Methodology.pdf

Hickman, R. C. (2011). First-semester academic performance as a predictor of fall-to-fall persistence: An application of classification tree analysis. *Journal of Applied Research in the Community College, 19*(1), 57-63.

Johnson, J. (2012). Big data + big analytics = big opportunity. *Financial Executive, 28*(6), 50-53.

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches.* Boston, MA: Pearson Education, Inc.

Jones, C. & Healing, G. (2010). Net generation students: Agency and choice and the new technologies. *Journal of Computer Assisted Learning, 26*(5), 344-356. doi:10.1111/j.3165-2729.2010.00370.x

Jones, W. A. (2011). Variation among academic disciplines: An update on analytical frameworks and research. *Journal of the Professoriate, 6*(1), 9-27.

Kalsner, L. (1991). Issues in college student retention. *Higher Education Extension Service Review, 3*(1), 3-9.

Maalouf, K.J. (1993). *The influence of late registration on academic outcomes and retention at a multi-campus community college.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Full Text database. (UMI No. 3534955)

Martella, R. C., Nelson, J. R., Morgan, R. L., & Marchand-Martella, N. E. (2013). *Understanding and interpreting educational research.* New York, NY: The Guilford Press.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60-68.

McClenney, K. M., & Waiwaiole, E. N. (2005). Student voices on student retention. *Community College Journal, 75*(5), 33-36.

McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction.* New York, NY: Longman.

Metz, G.W. (2002, October). *Challenges and changes to Tinto's persistence theory*. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED471529

Murray, C. (2010). Time, effort and care: The challenge or retaining students until they meet their academic goals. *Community College Journal, 80*(4), 34-35.

Oldfield, K. (2007). Humble and hopeful: Welcoming first-generation poor and working-class students to college. *About Campus, 11*(6), 2-12.

Parry, M. (2012, July 18). College degrees, designed by the numbers. *The Chronicle of Higher Education.* Retrieved from chronicle.com/article/College-Degrees-Designed-by/132945/

Pavolotsky, J. (2013). Privacy in the age of big data. *Business Lawyer, 69*(1), 217-225.

Pike, G., & Killian, T. (2001). Reported gains in student learning: Do academic disciplines make a difference? *Research in Higher Education, 42*(4), 429-454.

Royse, D., Thyer, B., Padgett, D., & Logan, T. K. (2006). *Program evaluation: An introduction.* Belmont, CA: Wadsworth.

Salter, D. W. (2003). Log-linear techniques for the analysis of categorical data: A demonstration with the Myers-Briggs Type Indicator. *Measurement and Evaluation in Counseling and Development, 36*(2), 106-121.

Schommer-Aikens, M., Duell, O., & Barker, S. (2003). Epistemological beliefs across domains using Biglan's classification of academic disciplines. *Research in Higher Education, 44*(3), 347-366.

Schroeder, C. (2013). Reframing retention strategy: A focus on process. *New Directions for Higher Education, 161*(1), 39-47.

Scott-Clayton, J., & Columbia University, (2011). The structure of student decision-making at community colleges. CCRC Brief. Number 49. *Community College Research Center, Columbia University.* Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED515132

Seymour-Campbell Student Success Act, California Code of Regulations, Title 5, Division 6 §§78210-78216(c). Senate Bill 1456 (2012).

Smith, V. C., Lange, A., & Huston, D. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks, 16*(3), 51-61. Retrieved from http://files.eric.ed.gov/fulltext/EJ982673.pdf

Spady, W. G. (1971, February). *The influence of major ambition resources on college aspirations and attainments: Toward a comprehensive model.* Paper presented at the 55th Annual Meeting of the American Educational Research Association, New York, NY. Retrieved from ERIC.

Summers, M. D. (2003). Attrition research at community colleges. *Community College Review, 30*(4), 64-84.

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

Tinto, V. (1987, November). *The principles of effective retention*. Presented at the Fall Conference of the Maryland College Personnel Association, Largo, MD. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED301267

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education, 68*(6), 599-623.

Tinto, V. (2004). Student retention and graduation: Facing the truth, living with the consequences. Occasional Paper 1. *The Pell Institute for the Study of Opportunity in Higher Education.* Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED519709

Tinto, V. (2008). When access is not enough. *The Carnegie Foundation for the Advancement of Teaching*. Retrieved from http://files.eric.ed.gov/fulltext/ED502271.pdf

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *Journal of Educational and Behavioral Statistics 39*(6), 478-501.

Trainor, S. (2015, 0ctober 20). How community colleges changed the whole idea of education in America. *Time*. Retrieved from http://time.com/4078143/community-college-history/

Trochim, W. (2006). *The research methods knowledge base.* Retrieved from http://www.socialresearchmethods.net/kb/index.php

Turner, P., & Thompson, E. (2014). College retention initiatives meeting the needs of millennial freshman students. *College Student Journal, 48*(1), 94-104.

U. S. Department of Commerce. (2013). *NIST/SEMATECH e-handbook of statistical methods.* Retrieved from http://www.itl.nist.gov/div898/handbook/

Varallo, S. M. (2008). Motherwork in academe: Intensive caring for the millennial student. *Women's Studies in Communication, 31*(2), 151-157.