The Relationship between Student Ratings and Student Retention

Isis Artze-Vega

University of Miami, isisartze@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
Artze-Vega, Isis, "The Relationship between Student Ratings and Student Retention" (2012). Open Access Dissertations. 785.
https://scholarlyrepository.miami.edu/oa_dissertations/785
THE RELATIONSHIP BETWEEN STUDENT RATINGS AND STUDENT RETENTION

By

Isis Artze-Vega

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Education

Coral Gables, Florida

May 2012
UNIVERSITY OF MIAMI

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

THE RELATIONSHIP BETWEEN STUDENT RATINGS AND STUDENT
RETENTION

Isis Artze-Vega

Approved:

Carol-Anne Phekoo, Ph.D.
Clinical Assistant Professor
of Higher Education

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Randall Penfield, Ph.D.
Professor of Research,
Measurement & Evaluation

Margaret Crosbie-Burnett, Ph.D.
Associate Professor of Educational
and Psychological Studies

Scott Ingold, Ed.D.
Student Functional Lead,
Campus Solutions ERP

Nicholas Myers, Ph.D.
Assistant Professor of Research,
Measurement & Evaluation

Molly Ott, Ph.D.
Lecturer of Higher Education
The relationship between students’ ratings of their instructors and student retention at the University of Miami was examined, extending a new thread of retention research which investigates the impact of students’ classroom experiences on their persistence.

The predictive ability of other factors known to impact retention (including students’ race/ethnicity, SAT scores, family income, and grade-point averages) was also examined, and these were controlled for to discern the unique contribution of student ratings to retention. The sample consisted of students who enrolled in this institution in 2002, 2003, and 2004, and the independent variables of interest were their ratings of their instructors (also known as student evaluations of teaching—SETs—and faculty evaluations). Overall ratings, as well as students’ perceptions of their instructors’ availability, presentation abilities, and their ability to generate interest in the course, were analyzed. Results indicate that students enrolled in highly-rated courses have enhanced odds of graduating within 6 years, that students with high GPAs benefit from being enrolled in classes with higher ratings, and that students’ total GPA is the strongest predictor of retention. Since faculty members are key determinants of students’ GPAs as well as of student ratings, implications for faculty development are offered.
For Abita and Mima, who paved the way,
and for Kamilah and Delilah, who have made me whole.
ACKNOWLEDGEMENTS

I must first express my gratitude to Dr. Carol-Anne Phekoo, who not only served as my dissertation committee co-chair, providing me with invaluable guidance during all stages of this difficult process, but who has also played the more instrumental roles of teacher, mentor, ally, and confidant during the last 7 years. Your steadfast dedication to my success and continuous encouragement were instrumental in my ability to complete this degree, and I will forever be grateful for your patience, insight, humor, and genuine concern. I was also very fortunate to have had in my corner a skillful and supportive dissertation committee: Dr. Randall Penfield, co-chair, I thank for his calming demeanor and pedagogical excellence; Dr. Margaret Crosbie-Burnett for her kindness and attention to detail; Dr. Scott Ingold for his candor and commitment to student and institutional success; Dr. Nick Myers for his statistical acumen and collaborative nature; and Dr. Molly Ott (last, but certainly not least) for providing conceptual and practical guidance which made an indelible impact on the quality of this dissertation. I would also like to acknowledge the fellow doctoral students who contributed directly and indirectly to this dissertation. First, the “early morning bagels club”: Chris Hartnett, Gary Markowitz, Carlos Reyes, and Cristy Sibila, whose camaraderie and support helped me complete (and even enjoy) this journey. Next, the statistical geniuses: Corinne Huggins and Allison Ames, whose assistance and instruction kept me sane and my methodology sound.

Finally, they say it takes a village to raise a child, so I cherish this opportunity to express my sincere thanks to my village; that is, everyone who helped me raise my daughters while mommy did her homework and worked on her “big project.” My Coral Gables crew, Chachis and Calixto, and my beloved in-laws, Maria Luisa and Luis,
provided countless hours of childcare (not to mention delicious meals). My lifelong pillars, my mom and dad, did all of this—and more. Everything good in my life was made possible by your unconditional love, boundless warmth, and unwavering support. Even so, as was plainly clear to me and to everyone around me, the most valuable player in this village was my larger-than-life husband, Sinuhé. For the innumerable sacrifices he made so that I could pursue and achieve this dream, for standing in for me at home and beyond the many times I was physically and/or mentally unavailable, and for showering me with love and putting a smile on my face day after day, I express my eternal gratitude.
# TABLE OF CONTENTS

| LIST OF FIGURES ............................................................................................................... vii |
| LIST OF TABLES ................................................................................................................... ix |
| LIST OF APPENDICES .......................................................................................................... x |

## Chapter

### 1 INTRODUCTION ........................................................................................................... 1  
Assumptions and Key Terms ................................................................................................. 6

### 2 LITERATURE REVIEW .............................................................................................. 9  
College Student Retention .................................................................................................. 9  
Retention Theories .............................................................................................................. 9  
Tinto’s Interactionalist Theory .............................................................................................. 9  
Adjustment of Tinto’s Theory for Minority Students .......................................................... 12  
Astin’s Theory of Involvement ............................................................................................. 14  
Bean’s Student Attrition model .......................................................................................... 15  
Terenzini and Reason’s Integrated Model .......................................................................... 16  
Retention Scholarship ........................................................................................................ 18  
Student Precollege Characteristics and Experiences ......................................................... 18  
The Organizational Context ................................................................................................. 20  
The Student Peer Environment and the Individual Student Experience .......................... 22  
Curricular and Classroom Experiences ............................................................................ 23  
Institutional Practice .......................................................................................................... 29  
Enrollment Management .................................................................................................... 30  
Student Ratings ................................................................................................................ 34  
Overview and Faculty Apprehensions .............................................................................. 34  
Psychometrics ................................................................................................................... 37  
Reliability ........................................................................................................................... 37  
Validity ................................................................................................................................ 38  
Evidence Based on Test Content ...................................................................................... 39  
Evidence Based on Internal Structure ............................................................................. 42  
Criterion-related Evidence ................................................................................................. 43  
Evidence Based on Response Processes .......................................................................... 47  
Evidence Based on the Consequences of Testing .............................................................. 48  
Rationale and Research Questions ..................................................................................... 50

### 3 METHOD .................................................................................................................... 54
Sample.......................................................................................................................... 54
Measures .......................................................................................................................... 55

Dependent Variables.................................................................................................... 55
Independent Variables ................................................................................................. 56
Covariates ....................................................................................................................... 57
Analyses .......................................................................................................................... 61
Assumptions of Logistic Regression ............................................................................. 65
Model Fit and Effect Size .............................................................................................. 66

4 RESULTS ...................................................................................................................... 68
Preliminary Analyses ..................................................................................................... 68
Tests of Model Fit and Baseline Interpretations ............................................................ 72
Research Questions 1 and 2: The Impact of GLOBAL ................................................... 74
Research Questions 3 and 4: The Impact of Pedagogical Behaviors ......................... 77
Interest ............................................................................................................................. 77
Availability ....................................................................................................................... 79
Presentation ...................................................................................................................... 81

5 SUMMARY, CONCLUSIONS, DISCUSSION, AND IMPLICATIONS .......... 84
Summary .......................................................................................................................... 84
Findings ............................................................................................................................. 87
Conclusions ...................................................................................................................... 88
Discussion ......................................................................................................................... 89
Practical Implications ..................................................................................................... 94
Implications for Ratings Data Interpretation ................................................................. 94
Implications for College Instruction .............................................................................. 95
Implications for Enrollment Management ..................................................................... 98
Recommendations for Future Research ....................................................................... 99
Limitations ....................................................................................................................... 99

REFERENCES ............................................................................................................... 104
LIST OF FIGURES

Figure 1. A Comprehensive Model of Influences on Student Learning and Persistence......................................................................................... 18
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Characteristics of an Excellent Teacher</td>
<td>41</td>
</tr>
<tr>
<td>Table 2</td>
<td>Descriptive Statistics for Ratings Variables</td>
<td>57</td>
</tr>
<tr>
<td>Table 3</td>
<td>Descriptive Statistics for Categorical Variables</td>
<td>60</td>
</tr>
<tr>
<td>Table 4</td>
<td>Descriptive Statistics for Continuous Variables</td>
<td>61</td>
</tr>
<tr>
<td>Table 5</td>
<td>Results of t-tests and Cohen’s d Values for Relationship between</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Continuous Independent Variables and Graduation within 6 Years</td>
<td></td>
</tr>
<tr>
<td>Table 6</td>
<td>Summary of Baseline Logistic Regression Model to Predict College</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Graduation</td>
<td></td>
</tr>
<tr>
<td>Table 7</td>
<td>Summary of Logistic Regression Model to Predict Impact of GLOBAL</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Student Ratings on Graduation</td>
<td></td>
</tr>
<tr>
<td>Table 8</td>
<td>Summary of Logistic Regression Model to Predict Impact of INTEREST</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Student Ratings on Graduation</td>
<td></td>
</tr>
<tr>
<td>Table 9</td>
<td>Summary of Logistic Regression Model to Predict Impact of</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>AVAILABILITY Student Ratings on Graduation</td>
<td></td>
</tr>
<tr>
<td>Table 10</td>
<td>Summary of Logistic Regression Model to Predict Impact of</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>PRESENTATION Student Ratings on Graduation</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendices</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>118</td>
</tr>
<tr>
<td>Intended and Unintended Consequences of Using Student Ratings Data</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>119</td>
</tr>
<tr>
<td>Table B1. Variable Definitions and Coding</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>120</td>
</tr>
<tr>
<td>Table C1. Number of Missing Cases per Variable</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>121</td>
</tr>
<tr>
<td>Table D1. Correlation Matrix among Continuous and Continuous Variables</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Table D2. Chi-square Values for Relationships among Categorical Independent Variables</td>
<td>122</td>
</tr>
</tbody>
</table>
Chapter One: Introduction

Perhaps no topic in American higher education has consumed scholars more than what has become known as the “departure puzzle,” a term coined by Braxton, Sullivan, and Johnson (1997) to describe the complex and pervasive phenomenon of college students’ departure from postsecondary educational institutions. For more than eighty years, this troubling occurrence has been the subject of empirical research (Braxton, 2000), resulting in “an ever more sophisticated understanding of the complex web of events that shape student leaving and persistence,” affirms Tinto (2006-7, p. 1). The fact remains, however, that efforts to improve retention have largely been ineffective, as demonstrated by unrelenting attrition rates (Reason, 2009): In 2010, two-thirds of students withdrew from college after the first year (ACT, 2011). One out of every two of the 18.3 million students enrolled in American colleges and universities in fall 2010 will not graduate, projects Yale (2010). Even in private research universities, where these statistics are more favorable, nearly 20% of undergraduates do not graduate within six years (Chronicle Almanac, 2011). In essence, during the last 50 years, our country has made great gains in expanding access to higher education, yet completion rates have remained virtually stagnant, and time-to-degree rates have risen considerably (Bowen, Chingos, & McPherson, 2009).

Even more troubling are the wide and enduring disparities in degree attainment related to race/ethnicity and socio-economic status (Bowen, Chingos, & McPherson, 2009). A recent study found that 60% of Whites but only 49% of Hispanic and 40% of African-American students who start college attain a bachelor’s degree six years later (Lynch & Engle, 2010). The college enrollment of Hispanics—the fastest growing U.S.
minority group—is projected to increase 38% between 2007 and 2018, more rapidly than any other group (Santiago, Kienzl, Sponsler, & Bowles, 2011). As the Obama Administration articulated and has been determined by a benchmarking analysis (Santiago et al., 2011), “it will be nearly impossible to reach our national [degree completion] goals without significantly increasing Latino college completion” (p. 5).

The “massive and continuing exodus from higher education,” as Tinto (1993) has described student attrition, has significant consequences for students because they cannot fully reap the occupational, monetary, or other benefits of higher education unless they attain a degree (p. 1). College completion has been linked to higher earnings, lower unemployment rates, greater civic and volunteer engagement, and reduced criminal involvement (Pascarella & Terenzini, 2005). From the perspective of colleges and universities as organizations, the stakes are likewise considerable: “These rates of departure negatively affect the stability of institutional enrollments, budgets, and the public perception of the quality of colleges and universities” (Braxton, Hirshy & McClendon, 2004, p.1). More specifically, Ackerman and Schibrowsky (2007-2008) estimated that on average, private, four-year colleges lose nearly $50,000 for each student that departs. Today, these challenges are intensified by an increasingly competitive postsecondary educational market as well as by ranking systems that use graduation rates as indicators of institutional quality and effectiveness (Kalsbeek & Hossler, 2010; Tinto, 2005). Kalsbeek and Hossler (2010) have therefore avowed that “the current level of degree attainment constitutes a crisis for America’s universities and erodes public confidence in higher education in general” (p. 5).
Coinciding with this retention predicament is a resurgent interest in college student learning. In September of 2006, a report released by then Secretary of Education Margaret Spellings’ Commission suggested sizable changes to the accreditation process and mandatory testing to measure college student gains, and in so doing, unapologetically launched a national movement toward holding colleges and universities accountable for student learning. Indeed, the problems identified in the report—including soaring college costs, signs that college students fail to learn key competencies and skills, and the lack of data with which to examine the outcomes of higher education—sparked national dialogue. As a result, demands for evidence now hail from nearly every sector of society: all levels of government, the general public, as well as students and their families (Webber & Boehmer, 2008). These parties will no longer accept exorbitant college costs without evidence of student learning. Most recently, the publication of Academically Adrift (Arum & Roksa, 2011) rekindled interest in the matter and made headlines with its research-based proclamation that many college students fail to make significant cognitive gains during their college years.

It is not surprising, then, to find that the topics of retention and college learning have become fixtures in the national public policy discourse. On the one hand, the U.S. federal government has historically played a limited role in higher education and provides only a fraction of all college and university revenues. On the other hand, federal expenditures for higher education far exceed those of states, industry, and other stakeholders in terms of direct aid to students and research and development funds (Gladieux, King, & Corrigan, 2005), and federal outlays amounted to roughly 48 billion in 2010 (NCES, 2010). This patronage is accompanied by increased calls for evidence of
student learning other than course grades (Middaugh, 2007). Kalsbeek and Hossler (2010) provide another rationale for the critical contemporary stance toward higher education: From a historical perspective, U.S. postsecondary education is now considered a “mature industry,” resulting in a public policy emphasis on accountability, quality control, and performance/productivity (as opposed to features previously stressed, such as growth).

As is true for any educational institution, the most coveted indicator of effectiveness for a college or university is empirical evidence of student learning. At this educational level, however, accountability is complicated by a variety of factors, including the abstract nature of many of the aims of a liberal education, the innate complexity of cognitive development, and the fact that learning has long been considered the purview of faculty members and thus distinct from the world of administration. With respect to this last point, Gaff (2007) describes the significant (and ironic) disconnect between authority and accountability that has long characterized academia:

It is generally agreed that the faculty, those with expert authority, should be the ones to make academic decisions rather than administrators or trustees, who have bureaucratic authority. Yet while the faculty are generally responsible for academic decisions, they are seldom held accountable either for student learning or for the fiscal results of their decisions. And while administrators are held accountable for student learning by accrediting agencies, they have no legitimate authority to intervene in the academic programs that are designed to produce student learning; while they are responsible for financial prudence, they again have little authority to ‘meddle’ in the curriculum or to alter academic decisions made by the faculty (p.6, italics his).

In essence, then, the general issue of accountability—perhaps more than any other feature of contemporary higher education—unites the academic and administrative domains, and
reminds us that the two must work together toward the common goals of student achievement and institutional well-being.

More specifically, Gaff’s (2007) excerpt brings up the contentious issue of faculty member liability, an increasingly prominent concern in public discourse. In fact, many of those who lambast colleges and universities point the finger at so-called incompetent instructors (Riley, 2011; Sykes, 1988). These parties demand that faculty members assume responsibility for student learning and provide evidence that students are receiving quality instruction (Hersh & Merrow, 2005). In the case of student retention, research has confirmed that faculty members play a significant role in the process (Pascarella & Terenzini, 2005), but again, it is usually administrators who orchestrate interventions and strategies aimed at retaining students (Tinto, 2006-7). The question then becomes, how can institutions certify the quality of their faculty members while respecting faculty autonomy?

One well-known practice that attempts to do so is the systematic collection of student feedback through end-of-course surveys, now ubiquitous on college campuses nationwide. Established in the U.S. in the mid-1920s, student ratings appear to have been the brainchild of Harvard students who sought to help their peers select courses, yet the practice of surveying students was quickly embraced as a means of obtaining data with which to assess teacher quality (Doyle, 1983; Weimer, 2010). In this regard, the use of student ratings can be said to represent “progress in the long journey to gain respect for teaching,” as instructor quality had previously been evaluated by hearsay (Weimer, 2010, p. 49). Scholars have been studying rating instruments and data ever since. There has been a resurgent interest in the topic due to the heightened concern with undergraduate
learning and accountability (Kahn, 1993; Ory, 2000). All the while, many—many faculty members, in particular—have argued virtually since the conception of this practice that data collected through student surveys are unreliable and invalid (Arreola, 2005; Ory, 2001).

The review that follows responds to these concerns by describing an extensive body of scholarship which demonstrates that students’ summative assessments of faculty can, in fact, be dependable measures of instructional quality (Arreola, 2005), and that student rating data are indispensable to those charged with institutional assessment. This study aimed to thus utilize student rating data to shed light on the student departure puzzle. It examined the extent to which taking highly-rated courses predicts student retention. My overarching hypothesis, then, was that students enrolled in highly-rated courses would be more likely to graduate within six years. Accordingly, I anticipated that my findings would lend credence to Braxton, Bray, and Berger’s (2000) pronouncement that “institutions that promote strong teaching skills…are likely to reap the twin benefits of increased student persistence and greater student learning” (p. 226).

Assumptions and Key Terms

Using a variety of terms to describe retention and its antithesis (including persistence, attainment, completion, graduation; attrition, dropout, student departure), researchers long presumed that every student’s ultimate goal was to secure a degree. More and more, however, scholars realize that college students are a diverse group with diverse objectives. Reason (2009) points out that “individual students define their goals,” which means that “a student may successfully persist without being retained to graduation” (p. 660) if, in fact, his/her goal did not mandate the completion of a degree.
Students may choose to voluntarily leave a college or university because they find their needs are not being met, have already gained whatever knowledge or skills they hoped to gain from college, and/or no longer think the return on the investment will be worthwhile. Although I agree with this perspective, this study nonetheless viewed graduation as the ultimate goal, both for the student and the institution, as this is presumably the goal of most University of Miami undergraduates.

Moreover, Reason (2009) critiques the fact that researchers use the terms “retention” and “persistence” interchangeably. I have heeded his suggestion that “retention” be used when focusing on institutional needs, and “persistence” when concerning student needs and behaviors. The term “departure” is used to describe the converse of retention, and “student attrition” will be defined herein as “the cessation of individual student membership in an institution of higher education” (Bean, 1980, p. 157). Since the primary outcome of interest is the University of Miami’s ability to retain its students, it will be referred to as “retention,” and retention-to-graduation (within six years) will be examined, as this is a commonly used benchmark institutionalized by the federal “Student-Right-to-Know and Campus Security Act” passed in 1990 (Selingo, 2012).

On the matter of student ratings of instruction, the study recognizes they are at best indirect measures of student learning yet rests on the assumption that they are a critical source of information representing the perspective of the very individuals we hope to retain and educate. This study uses the term “student ratings” as opposed to commonly used alternatives for three reasons: (a) There is little evidence to suggest that students’ responses are “a judgment of worth or value, usually after careful appraisal and
study,” as the term “student evaluations” suggests (Langbein & Snider, 1999, p. 457); (b) the use of “faculty evaluations” reflects the unfortunate fact that these ratings are often the only component of many faculty evaluation systems (Arreola, 2005); and (c) “course evaluation” is inaccurate since ratings primarily reflect students’ views of the instructor teaching the course, not of the course being taught (Marsh, 1987, p. 259).

In addition, this study focuses on within-classroom influences on student retention (i.e. student ratings) to work against the social class bias inherent in most early retention studies (Tinto, 2006-7). Since students from lower socio-economic classes are less likely to attend residential institutions or spend considerable time on campus, scholars’ prior emphasis on outside-of-class factors may have painted an incomplete picture of these students’ experiences. Although this does not appear to have been an intentional bias, I have explored students’ classroom experiences to attempt to enhance our understanding of all students’ persistence. Finally, this study affirms that faculty autonomy must be respected and that faculty members cannot be expected to improve the quality of their teaching without the availability of high-quality professional development opportunities (Arreola, 2005; Weimer, 2010). It is my hope that the University of Miami will use the findings of the study to make strategic decisions and, especially, to inform faculty development efforts.
Chapter Two: Literature Review

This chapter consists of two main parts: a review of conceptual frameworks and empirical studies on student retention, and a review of research on student ratings.

College Student Retention

Given that college student persistence constitutes an expansive body of knowledge and has been thoroughly reviewed by top scholars in the field of higher education (notably Pascarella & Terenzini, 2005; Reason, 2009, 2003; Tinto, 2005-6), the review that follows presents only the foundational concepts and most recent empirical studies pertinent to the study at hand. Also, although there is a large body of conceptual and empirical research on two-year colleges, this review focuses on four-year colleges.

Retention theories. The sections that follow describe four major conceptual frameworks that attempt to explain student retention. Although they all inform this study in various ways, the study design is most directly influenced by Tinto’s Interactionalist Theory and Reason and Terenzini’s comprehensive model, for reasons described below, with an emphasis on the category of retention factors that Reason and Terenzini refer to as “individual student experiences.” The study’s more-specific emphasis is on two subsets of this category: students’ classroom and curricular experiences, since the student ratings data that was utilized encompasses these two constructs.

Tinto’s Interactionalist Theory. Vincent Tinto’s (1975, 1987, 1993) now paradigmatic Interactionalist Theory of Student Departure submits that retention is a function of “the dispositions of individuals who enter higher education, …the character of their interactional experiences within the institution following entry, and…the external forces which sometimes influence their behavior within the institution” (1993, p. 37). In
describing student “dispositions,” Tinto posits that students’ (educational or occupational) intentions as well as their commitment (or motivation/drive) are inextricably tied to their persistence. In the case of “interactional experiences,” that is, the influence of events occurring within an institution that impact student experiences, Tinto’s work is informed by social anthropologist Arnold Van Gennep’s (1960) study of rites of passage. Van Gennep describes stages of separation, transition, and incorporation through which individuals pass in order to establish membership in traditional societies. Tinto argues that the manner in which students enter and become incorporated into a college community (particularly during the first year) parallels these rites of passage. In this respect, students must first separate or disassociate themselves with their past lives; then transition to college, becoming increasingly familiar with its norms and patterns; and finally become incorporated into the life of the college, adopting its customs as their own.

Tinto’s theory of departure also borrows from sociologist Emile Durkheim’s (1951) theory of suicide, one that classifies suicide into four types: altruistic (or morally desirable), anomic (during plague, war, for instance), fatalistic (in response to excessive regulation), and “egotistical.” Durkheim postulated that “egotistical suicide” was the result of an individual’s inability to become integrated and establish membership in a community, and he subdivided integration into social and intellectual. Applying Van Gennep’s stages and Durkheim’s concepts to the study of persistence, Tinto theorizes that college student departure mirrors this last form of suicide, as it reflects a student’s inability to pass through the stages of passage and become integrated into a college’s social and academic systems. More precisely, he finds that levels of integration are directly linked to a student’s decision to remain or depart, and that persistence is a
function of congruence between a student’s motivation and academic ability and an institution’s academic and social characteristics. To summarize, Tinto’s theory is that increased levels of integration strengthen students’ commitment to the institution and thereby increase the likelihood they will be retained.

Since both concepts are pivotal to this study, academic integration and social integration are worth expanding upon a bit. Influenced by students’ initial level of goal commitment, academic integration refers to students’ perceptions of their academic experiences and activities (Ishitani & DesJardins, 2002-3). Tinto (1993) describes this form of integration as membership in an institution’s academic systems, that which concerns students’ formal education: classrooms, laboratories, faculty, etc. Scholars have typically defined and measured academic integration as “a student’s estimation of their academic and intellectual development, grade point average, and students’ perception of faculty concern for teaching and student development” (Braxton, Milem, & Sullivan, 2000, p. 571). Grades are frequently utilized to measure academic integration since they are the most tangible evidence of rewards in the educational system (Ishitani & DesJardins, 2002-3).

Social integration, in turn, refers to students’ social involvement with peers and faculty. Tinto (1993) describes it as students’ membership in an institution’s social system, that which is centered on “the daily life and personal needs of the various members of the institution, especially the students,” and occurs mostly in residential halls, cafeterias, and other “meeting places” (p. 106-7). Voluntary student departures are thought to result from a lack of social support groups as well as a lack of congruence with peers, as this would result in an inability to become socially integrated (Tinto, 1987). In
addition, interacting with faculty members in and outside of the classroom (in office hours, before and after class, over dinner, etc.) is considered a crucial component of students’ social integration because instructors are socializing agents who convey institutional norms and values to students (Tinto, 1993).

Adjustment of Tinto’s theory for minority students. Although Tinto’s theory is widely respected and referenced, it is worth noting that several of its postulates have been contested by minority student scholars. In general, these scholars disapprove of the fact that Tinto’s model appears to place the responsibility for retention almost entirely in the hands of students, and they advocate instead for joint student-institutional responsibility for retention. Kuh and Love (2000) avow that alternative approaches to studying student departure are warranted for this very reason—that is, that Tinto’s model “is based on assumptions that understate institutional responsibility for creating hospitable learning conditions” and attributes “little or no responsibility to the institutions to modify their policies and practices to respond to the changing needs and characteristics of students” (p. 198). Others have challenged Tinto’s contention that financial difficulties and academic achievement are not significant factors in student attrition (Zurita, 2004-5).

Nevertheless, the most pronounced objection to Tinto’s theory appears to center on the notion that students must disassociate from their home community (a tenet of the separation stage) to effectively integrate into the college community (Guiffrida, 2006; Zurita, 2004-5). As Gonzalez (2000-1) puts it,

One…harmful consequence of applying Tinto’s model is that for racial and ethnic minority students to persist in college, they must first go through a ‘cultural suicide’ of their own culture in order to become socially integrated (or assimilated) into the dominant, Anglo, culture of their college or university. (p. 70)
Expanding on this point, Guiffrida (2006) notes that the separation stage of Tinto’s theory ignores bicultural integration, “the ability of minority students to succeed at college while being part of both the minority and majority cultures” (p. 452). For all of these reasons, scholars have begun to develop culturally-based perspectives on student departure.

One concept that has been proposed as an additional part of the psychosocial process involved with minority students’ adjustment and transition to college is known as “sense of belonging” (Hurtado & Carter, 1997; Johnson et al., 2007; Locks, Hurtado, Bowman, & Oseguera, 2008). In short, sense of belonging denotes the extent to which students feel “at home” in the college where they are enrolled (Meeuwisse, Severien, & Born, 2010). Also referred to as “perceived social cohesion,” sense of belonging is considered a key outcome of students’ experiences with social and academic integration on campus (Locks et al., 2008). In contrast to Tinto’s original interactionalist standpoint, the sense of belonging perspective views neither acculturation nor assimilation as necessary for student persistence.

Meeuwisse, Severien, and Born (2010) point out that a frequent theoretical underpinning used in “sense of belonging” studies is the concept of institutional *habitus*, which helps elucidate how this outcome relates to student retention. They refer to Berger’s (2000) explanation that “each campus is composed of students who generally share a common *habitus* which to some extent is congruent with the organizational *habitus* of that institution” (p. 531-2). According to this theory, students who share behaviors or are skillful in reading normative cues adapt more easily to the dominant peer group and are thus more likely to “fit in.” The general idea is that “the similarity of shared backgrounds, aspirations, and attitudes among students who constitute the
dominant majority on campus probably makes it easier for these students to adapt to campus life,” whereas adaptation is likely more challenging for students from underrepresented groups (Meeuwisse, Severien, & Born, 2010, p. 531-2). As pertains to retention, students who do not feel a strong sense of belonging—because of their social and cultural practices and/or tacit knowledge—may be more predisposed to withdraw from the institution (Meeuwisse, Severien, & Born, 2010).

**Astin’s Theory of Involvement.** Another conceptual framework for understanding retention, one similar to Tinto’s, is Alexander Astin’s (1975, 1999) theory of college student involvement. Astin’s fundamental contention is that the more involved students are in the academic and social life of an institution, the more likely they are to learn and persist. Based on a longitudinal and multi-institutional study of students who enrolled as freshmen in 1968 and were again surveyed four years later, Astin examined entering students’ personal background data, their experiences while enrolled, as well as the degree of fit between student and the college. He also developed the construct of “student involvement” (also known today as “student engagement”) to describe to the amount of physical and psychological energy a student devotes to the social and academic realms of the college experience.

Much like Tinto viewed grades as indicators of academic integration, Astin proposed that a high grade-point average signals “academic involvement,” based on the strong correlation between grades and persistence he identified in his study. Astin also found evidence that extracurricular activities, on-campus living, and on-campus employment were positively correlated with persistence, thereby forming the basis for his construct of social involvement, which he considers just as important to student retention.
as academic involvement. High levels of involvement have also been shown to be independent predictors of learning gains (Tinto, 1997). Finally, since Astin’s model of involvement resembles Tinto’s model of integration in the sense that they both have an academic and a social component, it seems necessary to explicitly differentiate academic and social involvement from academic and social integration: The former are indicators of physical and psychological energy devoted to the college experience, while the latter describe extent of membership in the college community. Therefore, involvement can be considered a precursor to integration (Milem & Berger, 1997). For instance, students who devote a great deal of energy to academics by tutoring their peers (a form of academic involvement) would likely feel that they are members of their college’s academic community and confident in their academic abilities (signs of academic integration).

**Bean’s Student Attrition Model.** John Bean’s (1978, 1982, 1983, 1985) Student Attrition Model constitutes the other dominant strand of retention research; the counterpoint, as it were, to Tinto and Astin’s theories. While Tinto and Astin’s approaches to exploring student departure are sociological in nature, Bean’s is primarily psychological. Using a worker turnover framework (a kind of attitude-behavior model), Bean viewed student attrition as analogous to employee turnover. He suggested that a number of student and institutional factors affect two intervening variables: satisfaction with college and intent to leave, the latter being a direct precursor to student attrition. Among the factors that influence student satisfaction in Bean’s model (all transposed from employee turnover theories) are routinization, centralization, development, and distributive justice.
Routinization signals “the degree to which the role of being a student is viewed as repetitive” (Bean, 1980, p. 159) and is predicted to negatively impact students’ satisfaction and thus increase their chances of dropout. Centralization refers to the extent to which students feel they are included in decision-making processes; development indicates the degree to which students feel they are developing as a result of being enrolled in the institution; and distributive justice is the sense that rewards and punishments are distributed fairly, i.e. based on effort. These three variables are predicted to positively impact students’ satisfaction and thus decrease their chances of dropout. To these, Bean added grades, practical value of education, the availability of desired courses, and membership in campus organizations. In 1987, Metzner and Bean extended the model because they felt the original version failed to adequately account for the experiences of nontraditional students, adding such factors as hours of employment, family responsibilities, and the opportunity to transfer to another college.¹

**Terenzini and Reason’s integrated model.** Tinto, Astin, and Bean’s theories are unquestionably the cornerstones of retention research and practice, yet Terenzini and Reason’s (2005) retention framework is also worth expounding as it is the most comprehensive depiction of student persistence to date. In the conclusion of their award-winning review of more than thirty years of research on the impacts of college on students, *How College Affects Students*, Pascarella and Terenzini (2005) pointed out what they consider the greatest weakness of student impact studies, namely that most of them

¹ There is reason to believe that this last point will become more salient in retention scholarship. The *Chronicle of Higher Education* (Brainard & Fuller, 2010) recently reported that, of nearly 1,400 four-year institutions analyzed, one-third reported lower graduation rates for the six-year period ending in 2008 than for the one ending in 2003. Among the chief reasons the authors cite is the fact that retention rates, as currently calculated, fail to take into account the experiences of students who transfer in or out of an institution.
have focused too narrowly on one or a handful of factors that impact retention at a time. The authors argued that the interconnectedness of the variables had not yet been adequately explored. Responding to this call to action, Terenzini and Reason (2005) developed an inclusive and integrated model for studying student outcomes, including student retention. In so doing, they incorporated “an emphasis on the organization’s effects on student outcomes” (which most other models had left out) as well as “internal organizational features such as policies affecting course sizes, promotion and tenure, or budgetary and staffing arrangements” (Reason, 2009, p. 662).

Terenzini and Reason’s (2005) framework includes four sets of constructs that encompass the wide array of variables shown by scholars to affect student persistence: (a) student precollege characteristics and experiences (including sociodemographic traits, academic preparation and performance, and student dispositions); (b) the organizational context, (c) the student peer environment; and (d) the individual student experience (a subset of the student peer environment), comprised of classroom experiences, out-of-class experiences, and curricular experiences (See Figure 1 below). The model may seem overwhelming and may appear to “paint a picture of futility” for those charged with retention given the magnitude and complexity of this framework, Reason (2009, p. 679) recognizes, yet there is a bright side, he affirms: “If… organizational contexts precede and influence the peer environment and individual student behaviors, institutional policies and practices are powerful levers for increasing engagement and persistence” (p. 679, emphasis added). In other words, he stresses the hopeful finding that colleges and universities can, in fact, improve their retention rates through institution-specific behaviors, an optimistic perspective that undergirds this study.
Retention scholarship. The four sets of constructs in Terenzini and Reason’s (2005) framework represent the four major categories of empirical research that have been conducted on the topic of college student retention.

Figure 1. A Comprehensive Model of Influences on Student Learning and Persistence

Student precollege characteristics and experiences. The precollege category consists primarily of sociodemographic traits, and there is no shortage of literature on the impact of gender, race, ethnicity, and socioeconomic status (SES) on retention. Although evidence of the effects of gender remains inconclusive (Astin, 1975; Reason 2001; Reason, 2003; Tinto, 1987), strong correlations have been found between students’ attrition rates and their race/ethnicity, SES, and income background. White and Asian

students persist at significantly higher rates than students from all other racial and ethnic 
groups (Astin, 1997; Braxton, Brier, & Hossler, 1988; Murtaugh et al., 1999). For 
instance, Oseguera (2005-6) reports four-year completion rates of 21, 23, and 38% for 
Mexican American, African American, and Asian American students, respectively. 
However, as Reason (2009) points out, after controlling for other important variables 
(SES and academic preparation, for instance), “racial differences disappear or are 
reversed,” suggesting that “differences in income or preparation, not race, might be at the 
root of differences in student persistence” (p. 663).

Indeed, a higher level of socioeconomic status has been shown repeatedly to 
positively affect retention (Pascarella & Chapman, 1983; Pascarella & Terenzini, 1991, 
2005; Walpole, 2003). ACT (2004) reported that students’ SES was second only to high 
school GPA as a predictor of college retention. The correlation between income 
background and retention is perhaps most vividly illustrated by the U.S. Department of 
Education’s (2005) finding that the lowest-achieving students from high SES families 
complete college as frequently as the highest-achieving students from low SES families 
(Fox, Connolly, & Snyder, 2005). The importance of parents’ educational background 
(years of schooling), in turn, has been corroborated by several studies (Pascarella & 
Chapman, 1983; Pascarella & Terenzini, 1978). This variable has also been examined 
indirectly in research on first-generation college students (the first in their family to 
attend college), which have found a higher risk of departure among this group of students 
(Ishitani, 2006; Nuñez & Cuccaro-Alamin, 1998). With respect to academic preparation 
and performance, high school grade-point-average and standardized test scores are

---

3 Student achievement in this study refers to achievement in 8th grade math.
known to be significant predictors of retention (Astin, Korn, & Green, 1987; Tross Harper, Osher, & Kneidinger, 2000), with high-school GPA considered the strongest precollege predictors of retention (ACT, 2007).

Student dispositions—the other type of pre-college characteristics in Terenzini and Reason’s (2005) framework—have been examined much less frequently than sociodemographic traits (Reason, 2009). As described above, Tinto factored students’ (educational or occupational) intentions as well as their commitment (or motivation/drive) into his model, suggesting that high goals and high motivation levels lead to increased persistence. Robbins et al.’s (2004) meta-analysis corroborated this claim, finding that academic goals, academic self-efficacy, and students’ sense of academic skills were all associated with their persistence. Some empirical data even suggest that perceptions about academic success may bear more on students’ academic success than high school GPA or standardized test scores (Brown et al., 2008). Research by ACT (2004, 2007) provides additional evidence that high levels of academic motivation, self-discipline, and self-confidence increase students’ odds of persisting through the third year of college.

The organizational context. This construct examines the college environment, including both the structural-demographic characteristics of an institution and its organizational behaviors. Structural-demographic studies have explored the relationship between such factors as size, curricular mission, admissions selectivity, and source of support (public vs. private) on student retention. Overall, research findings on the impact of size, mission, and source of support are mixed (Reason, 2009), leading Terenzini and Reason (2005) to conclude that what organizations do—their actions, policies, values, etc.—have a greater influence on student outcomes than what they are (their institutional
type, size, location, etc.). One example of what institutions can do to promote student persistence is award need-based grants (Porter & Barberini, 1989), a practice which will become increasingly necessary as the enrollment of students from low SES families continues to rise (Keller, 2001). It should also be noted, however, that evidence of the overall impact of financial aid on retention is mixed (DesJardins et al., 2002, Knight & Arnold, 2000).

The only two structural-demographic features shown to significantly impact student retention are (a) where students begin their college career, and (b) institutional selectivity. While attending historically black colleges or women’s institutions first is associated with increased persistence for African-American and women students, beginning a college career at a two-year college has the opposite effect (Long & Kurlaender, 2008; Pascarella & Terenzini, 2005). As pertains to selectivity, students who attend highly selective schools have higher rates of persistence (Reason, 2009). Then again, since many of the same variables described above (high school GPA, standardized test scores, etc.) are associated with selectivity, the positive relationship between selectivity and persistence is to be expected. Another view is that selective institutions have peer environments that promote persistence (Pascarella & Terenzini, 2005).

Studies of organizational behavior dimensions have used two disparate approaches: The first is to categorize colleges and universities into broad organizational types, including collegial, political, bureaucratic, and symbolic institutions. In essence, collegial institutions are communities of semi-autonomous individuals and units working collaboratively toward common goals; political institutions are driven by power, bargaining, and negotiation; bureaucratic institutions are business-like entities in which
rules and regulations are paramount; and symbolic institutions resemble houses of worship in that symbols, stories, and myths carry more weight than actual occurrences (Birnbaum, 1989; Bolman & Deal, 2003). Institutions characterized as collegial and symbolic have been found to have higher retention rates, while the converse was true in those appearing more political and bureaucratic (Reason, 2009). The second approach has been to examine institutional expenditures under the premise that spending reveals an institution’s priorities. Gansemer-Topf and Schuh (2006) found, for instance, that institutions which allocated increased spending on such areas as legal services and public relations exhibited lower rates of student retention, while expenditures for instruction were significantly and positively associated with first-year student persistence and six-year graduation rates for all institutions.

**The student peer environment and the individual student experience.** College students have been shown to be considerably influenced by their peers (Pascarella & Terenzini, 2005), so it may seem intuitive that the student peer environment influences persistence. In fact, the student peer environment facilitates and/or inhibits the process of social integration which has been empirically linked to institutional commitment and subsequent student persistence (Braxton & Lee, 2005). Described as the “the system of dominant and normative values, beliefs, attitudes, and expectations that characterize a campus’ student body” (Reason, 2009, p. 670), the student peer environment affects an individual’s ability to become integrated into the institution and/or the extent to which a student feels a sense of belonging there. One key part of the peer environment is a campus’ racial climate. Museus, Nichols, and Lambert (2008) found that campus racial climate indirectly affects persistence through its influence on goal commitment, social
and academic involvement, and institutional commitment. Similarly, Johnson et al. (2007) examined sense of belonging among first-year students from different racial/ethnic groups and found that positive perceptions of the campus racial climate were significantly related to students’ sense of belonging among all groups except Hispanics.

Research studies within the category of the individual student experience have examined students’ out-of-class experiences, curricular experiences, as well as (to a lesser extent) classroom experiences. Out-of-class experiences are associated with students’ level of engagement or involvement, which (as mentioned) has been shown to influence student persistence. More specifically, residing on campus significantly increased a student’s chances of attaining a degree (Astin, 1977; Chickering, 1974); participation in educationally purposeful activities (such as communicating with an instructor via email and working with classmates outside of class on an assignment) is positively related to first-to-second year persistence (Kuh et al., 2008); and involvement in student groups and organizations has been shown to be indirect and mixed (Pascarella & Terenzini, 2005). In the case of curricular experiences, certain academic majors (science, technology, education, and math, i.e., STEM fields, for instance) have been linked to higher degree completion rates (Pascarella & Terenzini, 2005), and enrollment in first-year seminars (such as freshman success courses) has likewise been associated with increased first-to-second year persistence (Hunter & Linder, 2005).

**Curricular and classroom experiences.** Not surprisingly since it is a prime signal of students’ academic integration, students’ college GPA has been found to be the most powerful in-college predictor of retention (Murtaugh, Burns, & Schuster, 1999; Pascarella & Chapman, 1983; Pascarella & Terenzini, 1978). As Tinto (2006-7) has
explained, student involvement in the classroom is important to student retention
“because the classroom is, for many students, the one place, perhaps only place, where
they meet each other and the faculty” (p. 4). Echoing this point, Reason (2009) indicates
that “the classroom may be the only context regularly inhabited by every student,”
particularly on campuses with numerous commuter students (p. 678). Attempts to
increase student engagement in academic settings is not only in keeping with the overall
mission of higher education, Reason adds, “but it [also] becomes imperative when
classrooms become the only organizational context students experience” (p. 678). Yet
only recently have scholars begun to recognize and examine the role of the classroom in
the student departure puzzle (Braxton, Milem, & Sullivan, 2000). In fact, in a
comprehensive review of persistence research, Reason (2009) notes that he found
“surprisingly little research exploring the connection between students’ in-class learning
experiences and persistence” (p. 673).

Perhaps the most noteworthy exceptions are studies of what are known as
“curricular interventions.” Early research on the effects of learning communities (one
curricular intervention in which a cohort of students takes several courses together) on
student involvement and retention was able to establish a clear link between classroom
practice, heightened involvement, and student persistence (Tinto, 1997). Reflecting on
the qualitative and quantitative results of his study of a learning community, Tinto (1997)
inferred that the initiative was able to increase students’ likelihood of persisting because
it enabled students to develop a support network, a community within the larger college
community, one that helped students bond to the institution. More recently, Keup (2005-
6) used a national dataset and follow-up surveys of freshman experiences, and found
positive relationships between three curricular interventions—first-year seminars, service-learning, and learning communities—and the longitudinal process of first-to-second year persistence.

Despite these studies of curricular programs and empirical evidence of their influence on persistence, examinations of students’ classroom experiences and interactions with faculty members are still scarce. It is worth noting that Pascarella and Terenzini (2005) found evidence that students make a number of cognitive gains during their college years, including gains in the ability to think critically and in post-formal reasoning. In 2000, Tinto argued that “we must reconstruct our theoretical models to include not only the classroom, but also the faculty and, in turn, pedagogy in our discussions of student persistence” (p. 81). In an impassioned discourse, he asserted that faculty members must be brought back into retention theories not only because contact with faculty inside and outside the classroom serves to directly shape learning and persistence, but also because their actions, framed by pedagogical assumptions and teaching skills, shape the nature of classroom communities and influence the degree and manner in which students become involved in learning in and beyond those communities. (p. 90)

In other words, he surmises that faculty members exert more influence on student retention than has been accounted for by researchers. Only a few years later, Tinto (2006-7) bemoaned a related problem and one all too common in academia: the disconnect between practice and research. The practice of trying to retain students does not reflect the “widely accepted notion that the actions of the faculty, especially in the classroom, are key to institutional efforts to enhance student retention,” he wrote (p. 5).

One research study corroborating Tinto’s assertion examined the impact of specific teaching practices on student retention. Braxton, Milem, and Sullivan (2000)
hypothesized that active learning\(^4\) positively impacts students’ social integration, rationalizing that the additional classroom interaction with peers and faculty inherent in active learning helps students develop a peer group. They surveyed first-year students at a highly selective, private research university and found that “faculty classroom behaviors in general and active learning in particular may constitute an empirically reliable source of information on social integration, subsequent institutional commitment, and departure decisions” (p. 582). The authors concluded more generally that faculty members’ classroom behaviors influence the student departure process.

Other scholars narrowed the scope further to investigate not what occurs within classrooms and what faculty members actually do (as in an observational study, for instance) but rather students’ perceptions of instructors and the extent to which students’ perceptions are associated with their degree completion. Noting the dearth of retention research on the role of faculty members in this process, Lundquist, Spalding, and Landrum (2002) investigated the impact of faculty attitudes and behaviors on students’ thoughts about leaving the university. They found that three issues significantly impacted students’ thoughts about departing: “faculty members being supportive of student needs, returning phone calls and e-mails in a timely fashion, and being approachable” (p.1).

In turn, Braxton, Bray, and Berger (2000) rationalized that if collaborative learning influences student departure decisions, so too might students’ perceptions of faculty teaching skills. “Students who take courses from faculty members who frequently exhibit [strong teaching] skills may be more likely to invest the psychological energy

\(^4\) Defined in the study as “any class activity that ‘involves students in doing things and thinking about the things they are doing’”—including discussions, debates, role-playing, and asking probing questions in exams.
necessary to establish membership in the social communities of their college or university,” (p. 216) they theorized. Otherwise put, the authors conjectured that good teaching affects students’ levels of social integration and may thus be a precursor to student persistence.

Braxton and his colleagues (2000) focused their examination on three faculty behaviors known to positively influence student course achievement: organization, preparation, and clarity. Using path analysis to test their hypothesis as it applies to first-year students and their intent to return (a proxy measure for college student departure), they found strong general support for their model. Two measures of student perceptions about faculty teaching skills (organization/preparation and instructional skill/clarity) were found to be positively related to social integration, subsequent institutional commitment, and intent to re-enroll. It seems worth mentioning that the survey items used to gauge students’ views of their instructors mirror items found in many end-of-course ratings (for example, “Course requirements are clearly explained”). Also, since Braxton, Bray, and Berger (2000) concede that “the influence of faculty teaching skills on student departure may differ for more advanced students than for first-year students” (p. 226), they encourage additional studies not limited to first-year students.

Taking what they called the “next logical step” in this line of inquiry, Pascarella, Seifert, and Whitt (2008) analyzed data from a single large research university to test whether Braxton, Bray, and Berger’s (2000) findings are upheld when one looks at actual student reenrollment (in lieu of “intent to reenroll”). They also created direct and indirect models of the possible relationship between student background characteristics, college experiences, and classroom instruction, factoring in the effect of students’ college grades
and their “educational satisfaction.” To measure this last point—that is, satisfaction with the overall experience of college—students were asked to “evaluate [their] entire educational experience at [the] institution” on a four-point scale. Pascarella, Seifert, and Whitt (2008) found not only that students’ actual persistence was again tied to their exposure to organized and clear instruction, but also that the positive impact of the exposure was mediated through higher levels of student satisfaction with their education.

To test the “robustness” of these findings, as Pascarella, Salisbury, and Blaich (2011) put it, they analyzed a sample of first-year students attending a variety of institutional types (research, liberal arts, community college). They found that students’ perceptions of effective instruction significantly increased their likelihood of persisting and that this indirect effect appears also to be mediated through student satisfaction with the quality of their educational experience. In short, “exposure to organized and clear instruction enhances student satisfaction with the overall college experience, which, in turn, increases the probability of a student’s re-enrolling at an institution for the second year of college” (Pascarella, Salisbury, & Blaich, 2011, p. 17). From an institutional standpoint, this finding is intriguing and important, the authors conclude, because the instructor qualities examined are not innate pedagogical skills or personal propensities; they are learnable instructional behaviors. In this regard, the findings suggest that investing in faculty development programs designed to enhance teaching “may return significant dividends in terms of increased student persistence at a range of different institutional types” (Pascarella, Salisbury, & Blaich, 2011, p. 18).

As in the research conducted by Braxton, Bray, and Berger (2000) described above, the items used by Pascarella and his colleagues (2011) to measure students’
perspectives on their instructors closely resemble end-of-course ratings questions, yet the methodology of all three studies differs from student ratings in several notable ways: First, students were only asked about a handful of instructor qualities. Organization and clarity are important instructional characteristics, but others such as an instructor’s availability or ability to encourage students are also associated with positive outcomes (Pascarella & Terenzini, 2005). Second, students were asked to indicate frequency; in other words, how often they have experienced each instructional behavior (For instance, how often have they seen “teachers give clear explanations”). Instead, rating surveys tend to use Likert scales with which students are likely more familiar. Third, students in these studies were asked to assess more than one faculty member at a time. That is, they were asked to take into consideration “all of the teachers with whom [they’ve] interacted.” It seems likely that when asked about individual instructors—as in end-of-course ratings—students’ responses are more precise and thus more useful for administrative decision-making and faculty development.

**Institutional practice.** To turn now to actual retention efforts on college campuses across America, it should be noted that initially, the responsibility for student retention was assigned to student affairs professionals. As Tinto (2006-7) pointed out, most retention efforts existed outside of, or tacked onto mainstream academic life, and (as he also recognized) this is still the case in many institutions. Take, for instance, Seidman’s (2005) synthesis of retention activities in U.S. colleges and universities:
To retain students, colleges have provided programs for the economically disadvantaged, programs for underrepresented students, programs and services for students with disabilities, women, and older adults reentering college or beginning college for the first time. Counseling programs have been strengthened to try to meet the needs of students. Job and career centers have been established to help students decide on career options and to provide a place for potential employers to meet students. (p. xii)

This list of programs and services implicitly corroborates the fact that most retention efforts have been situated outside of institutions’ educational core. At the same time, however, it has become an administrative mantra that student retention is “everybody’s business” (Tinto 2006-7), not simply that of student affairs professionals, a notion very much in keeping with the holistic view of retention utilized in enrollment management.

**Enrollment management.** Today, it is increasingly the case that colleges and universities have charged an enrollment management team with retention (Braxton, Bray, & Berger, 2000), and this is true of the University of Miami, such that an overview of enrollment management is in order. The term “enrollment management” was coined in the mid-1970s by Boston College administrators who realized that due to demographic changes (and forecasts) and the sheer number of postsecondary options available to students, they would have to do something purposeful to remain competitive and viable (Henderson, 2001; Hossler & Kalsbeek, 2008). Early enrollment management practices were therefore centered on recruitment, marketing, admissions, and financial aid. As early as the late 1980s, however, enrollment management started becoming much more encompassing, evolving into “a comprehensive process that touches every aspect of an institution and seeks not just to attract and enroll learners, but to retain them for the duration of their educational experience and beyond” (Bischoff, 2007, p. 2).
Hossler and Bean (1990) defined enrollment management as “an organizational concept and a systematic set of activities designed to enable educational institutions to exert more influence over their student enrollments” (p. 5). The main point here is what is known as intentionality: An institution need not sit passively and be at the mercy of demographic, regional, or other forces that influence its student enrollments. Instead, it can be proactive to maximize its chances of success. Hossler and Bean add that “organized by strategic planning and supported by institutional research, enrollment management activities concern student college choice, transition to college, student attrition and retention, and student outcomes” (p. 5). This second half of their definition indirectly explains why the term “strategic enrollment management” (SEM) is often used interchangeably with “enrollment management,” namely, because strategic planning is an integral part of the process. Although the authors touch on this point as well, it is worth stressing that another main feature of enrollment management is that it is a data-driven, research-dependent process (Hossler & Kalsbeek, 2008). The use of student ratings data herein is therefore very much in keeping with the tenets of enrollment management.

Contemporary enrollment managers continue to devote considerable time and energy to marketing, recruitment, and other efforts, but student retention has unquestionably become one their primary concerns (Kalsbeek & Hossler, 2010). By and large, this is because the government and the general public are preoccupied with graduation rates which, as mentioned earlier, are commonly used as a proxy for institutional effectiveness and weighted heavily in national and international comparisons of colleges and universities (Kalsbeek & Hossler, 2010). By the same token, the current state of higher education is one of diminishing resources which makes retaining students
an essential goal of enrollment management, leading Kezar (2004) to predict that the role of enrollment management on campuses will become increasingly prominent.

Interestingly, although not explicitly discussing enrollment management at the time, Astin (1975) offered two compelling reasons—one financial and one educational—why retention efforts may be even more fruitful than recruitment efforts:

In four-year institutions, any change that deters students from dropping out can affect three classes of students at once, whereas any change in recruiting practices can affect only one class in a given year. From this viewpoint, investing resources to prevent dropping out may be more ‘cost effective’ than applying the same resources to more vigorous recruitment. More important from an educational standpoint, changes that help students complete college represent a real service to them, whereas successful recruiting efforts may simply change students’ choice of institution. (p. 2)

Astin’s point that focusing on retention is more aligned with an educational institution’s raison d’être than recruitment resonates with a new trend in enrollment management literature, one that encourages situating enrollment management in an academic context.

One the longest standing enrollment management quandaries has been the structural question of what functions or units should form part of an institution’s EM efforts. These have included admissions, financial aid, orientation, academic advising, and various others, but have not typically included an institution’s academic realm, considered the exclusive domain of faculty members. However, there has been a “significant shift” over the past five years or so to stop dwelling on the issue of structure and “to refocus enrollment management through the academic lens” (Bischoff, 2007, p. i). Henderson (2005) argued that enrollment management “cannot succeed unless it is part of the academic fabric of the institution” because only when we look at enrollments through an academic lens, can EM “touch every aspect of institutional function and
In other words, enrollment management isolated from academic is inherently incomplete. David H. Kalsbeek, a top enrollment management practitioner and scholar, concurs. When asked what he considers the most promising new insights and understandings about retention, he (2010) affirmed that “improving an institution’s retention and degree completion rates by any substantial margin generally requires improving students’ academic experiences,” including new approaches to pedagogical reform and enhanced classroom instruction (p. 9). Enrollment managers may not have ventured into college classrooms in the past, but it seems clear that the pressing issue of student retention is moving them in this direction.

Critics of enrollment management would likely interject at this point that the practice of enrollment management reflects and has perhaps even contributed to everything wrong with academia today: empty competition based on college rankings, access and attainment gaps, and the overall corporatization of higher education. They would probably also be quite suspicious of the encroachment of enrollment management into an institution’s academic affairs. Although I too avow that faculty autonomy must be protected, I would respond to these critiques in much the same way Hossler and Kalsbeek (2008) have done; that is to concede that enrollment management may be associated with some of these deleterious changes in higher education, but to also resolutely state that it is by no means the cause. To be sure, enrollment management tools can be used in “the arms race” in American higher education, but they can and should likewise be used toward noble goals such as access and equity. The fact of the matter is that in today’s competitive landscape of postsecondary education, “the stakes are too high, the consequences too great not to be as deliberate, as knowledgeable, as intentional, and as
effective as possible in pursuit of an institution’s desired strategic enrollment goals” (Hossler & Kalsbeek, 2008, p. 4).

In essence, the allocation of resources and personnel on campuses nationwide and the extensive history of retention theory and research all verify the significance of this issue to stakeholders in higher education. This body of knowledge also confirms that student persistence is a complex phenomenon involving a multitude of variables. With respect to the study at hand, it is worth reiterating that Tinto’s theory of retention describes student integration into an institution’s social and academic realms as crucial to student persistence; that a current line of inquiry focusing on what occurs within classrooms suggests that students’ appraisals of instructors’ teaching practices can influence student persistence; and that the widespread practice of enrollment management is beginning to recognize the importance of students’ academic experiences to their overall success.

**Student Ratings**

**Overview and Faculty Apprehensions**

Student ratings of faculty have typically served four purposes: (a) to provide feedback to instructors, (b) to make administrative decisions about tenure and promotion, (c) to help students choose courses, and (d) for research on postsecondary teaching (Nasser & Hagtvet, 2006). Already widely used across the country as a systematic assessment of instruction, Feldman (2007) foresees that the use of ratings is likely to increase now that colleges and universities are beginning to prioritize good teaching and learner-centeredness. Indeed, student ratings programs intend to improve teaching at two levels: the institutional level and individual instructor level, as Kulik (2001) has
articulated. “They may influence an institution’s hiring decisions, merit increases, promotion and tenure decisions, and course assignments. Ratings may thus influence who teaches at a college, what courses they teach, and how much attention faculty members give to teaching,” he writes. “Rating results [also] give teachers information that they may use when trying to improve their own teaching,” recognizes Kulik (p. 15). Weimer (2010) echoes this last point, noting “the [enormous] potential for faculty to grow and develop as teachers based on feedback provided by students” (p. 51).

In terms of sheer volume, research on the topic of student ratings is comparable to that on college student retention. Ratings expert Herbert W. Marsh (1987) noted in his 135-page monograph on the subject that the study of student evaluations was “one of the most frequently emphasized in American educational research” (p. 260). This has continued to be the case. More interesting is Marsh’s (1987) suggestion that this “considerable base of research” is the result of many faculty members’ apprehension and uneasiness with them, despite the extensive empirical work demonstrating their integrity. In other words, Marsh speculated that faculty members are so convinced ratings provide an inaccurate picture of their teaching abilities that they become compelled to discredit these data by doing what many of them do best—conducting research. Others have proposed that instructors view these surveys as a control mechanism that inconspicuously infringes on their academic freedom (Haskell, 1997). Most ratings scholars share McKeachie’s (1997) view that faculty members’ critiques stem from the all-too-common misuse of ratings data in personnel decisions, especially high-stakes ones involving promotions, merit pay, and contract renewal.
The skepticism that surrounds student ratings has been attributed also to the fact that so many instruments used by U.S. colleges and universities are “home-made”—that is, created by the institution or department without conforming to research-based standards or being tested psychometrically (Arreola, 2005). This has led to pervasive anecdotes and myths about their reliability, validity, and utility (Arreola, 2005). Two of the best-known legends are that ratings are a personality contest and that instructors can guarantee high ratings by giving students high grades. Arreola (2005) contended that these myths are so entrenched in academia that they are considered “common knowledge.” This being the case, they “far overshadow the ‘truth’ concerning student ratings and other faculty evaluation tools” and therefore represent a sizable obstacle for anyone wanting to convince administrators and especially members of the professoriate that the data are, or can be, valid (p. 30-31).

Perhaps most germane to this study is Marsh’s (1987) notion that the myths and misconceptions are perpetuated by faculty members who feel threatened by student ratings because of their limited formal training in teaching. As Jones (2008) explains,

At best, they have received a course in pedagogy specific to their discipline as a requirement of their professional or doctoral degree program. At worst, which unfortunately is frequently the case, they have been trained only in their disciplines—as scientists and anthropologists and accountants—and are thrown into their classrooms to rely only on their experience as students to inform their practice as teachers. (p. 93)

Faculty aversion to student ratings is thus understandable, especially when these data are used to make the kinds of high-stakes decisions described above. Any effort to expand the use of these ratings (including the use in supporting retention proposed in this study), will therefore need to reassure faculty members that the instruments being used have
been tested psychometrically and conform to standards, attempt to dispel University-wide misconceptions about ratings data, and emphasize that the proposed use of ratings data is not for high-stakes administrative decisions, but rather meant to bolster professional development.

**Psychometrics**

**Reliability.** The reliability of student ratings refers to the consistency and dependability of the data secured through their administration (Braskamp & Ory, 1994). As Braskamp and Ory (1994) paraphrase the main questions of reliability, “Is an observation made at one point in time likely to be similar to an observation made at another? Are observations made by two different faculty members, students, or administrators likely to be similar?” (p. 91). Student ratings experts highlight three primary ways to determine the reliability of these data: by conducting item analyses (i.e., correlations among responses to different items intended to measure the same aspect of effective teaching), considering the stability of ratings over time, and examining inter-rater agreement among students within the same course (Centra, 1993; Marsh, 1987).

Others have argued that the last approach—inter-rater agreement—is the most appropriate because (given a sufficient number of students) the reliability of class-average student ratings is quite high (Marsh, 1987; Marsh & Roche, 1997). For instance, the estimated reliability of one of the most widely used ratings instruments, the Students’ Evaluation of Educational Quality (SEEQ), is about .95 for the average response from 50 students and .90 from 25 students (Marsh, 1987). Others have reported reliability coefficients of about .90 for similar, well-developed instruments (Arubayi, 1987; Costin et al., 1971). At the same time, however, scholars and practitioners alike stress the fact
that reliability is a relative concept: The reliability of student ratings data is contingent upon the use of carefully constructed forms and established procedures for their administration (Aleamoni, 1999; Arreola, 2005; Braskamp & Ory, 1994). Although this may seem self-evident, “the actual day-to-day decisions of those who use ratings are too frequently uninformed by the research literature that professional ratings practitioners hold as their canon,” maintain Theall and Franklin (2000, p. 96), such that reliable student ratings data may be attainable, but reliability cannot be taken for granted.

**Validity.** The issue of practice becomes even more relevant when discussing the validity of these data, a considerably more complex concept than reliability. One reason for this complexity is that popular discourse on faculty evaluation and ratings data is muddled by semantics. The large majority of those discussing “validity” are using the dictionary definition of the term (logically correct, appropriate, etc.) as opposed to the psychometric perspective which refers to “the effectiveness of the test in representing, describing, or predicting the attribute” of interest, suggests Arreola (2005, p. 32). Since faculty members and academic administrators tend to use the colloquial, inherently subjective definition, he explains, it is not difficult to see why the validity of ratings data continues to be questioned and the anecdotes about their lack of validity perpetuated.

With respect to empirical research, the validity of student ratings was “severely questioned” in the 1970s, “by the early 1980s… most [experts] viewed student ratings as valid and worthy of widespread use” (Greenwald, 1997, p. 1182). It is important to note, however, that the prevalent meaning of validity at the time differs considerably from that utilized today. Prior to 1999, most used the term validity to refer to “the extent to which a test produced information that was useful for a specific purpose” and categorized validity
into specific types: content validity, criterion-related validity, and construct validity.

When the Standards for Educational and Psychological Testing (Standards, American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999) were revised in 1999, the “trinity” categorization was abandoned, and a broader, more holistic view of validity was advanced. According to the 1999 Standards,

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation therefore involved accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each included interpretation must be validated. (AERA, APA, & NCME, 1999, p. 9)

The 1999 Standards also describes five types of validity evidence: (a) evidence based on test content, (b) evidence based on internal structure, (c) criterion-related evidence, (d) evidence based on response processes, and (e) evidence based on consequences of testing. This validation framework therefore provides a useful lens through which to examine research findings on the validity of student ratings.

Evidence based on test content. Much like the notion of “content validity” signaled “the faithfulness with which [a] test represents or reproduces an area of knowledge” (Arreola, 2005, p. 32), evidence based on test content gauges the extent to which the subject matter of the measure represents a content domain. As Ory and Ryan (2001) clarify the question at hand, “Is there a relationship between the contents of the student ratings form and the construct intended to be measured?” (p. 32). One of the most commonly used approaches for validation based on content is conducting a content
review, a process during which a panel of experts examines items and rates the extent to which they correspond to the intended trait(s) to be measured. As pertains to student rating forms, it appears validity evidence based on test content has been obtained informally in most cases; that is, survey items have been chosen by individuals who believe they reflect important dimensions of college teaching (Aleamoni, 1999).

The major obstacle that comes to light when one attempts to collect evidence based on content aspects is that “to make valid inferences about student ratings of instruction, the rating items must be relevant to and representative of the processes, strategies, and knowledge domain of teaching quality” (Ory & Ryan, 2001)—yet no one can agree entirely on what “teaching quality” at the college level entails. This notable absence of a clear definition of “excellent teacher” in the academy presents the biggest hurdle in the whole of faculty evaluation, argues Arreola (2005). In an attempt to offset this deficiency, Arreola (2005) proposes four characteristics of an excellent teacher “based on a career-long assimilation of research literature and professional experience” (p. 33): content expertise, affective traits/skills, performance skills, and cognitive skills (described in Table 1 below).

Chickering and Gamson’s (1987) “Seven Principles for Good Practice in Undergraduate Education,” one of the cornerstones of postsecondary teaching and learning literature, offers another straightforward and time-tested point of reference that may facilitate the process of validation concerning content. Based on their examination of 50 years of research on teaching and learning, Chickering and Gamson (1987) wrote that good practice in undergraduate education: (1) encourages contacts between students and faculty, (2) develops reciprocity and cooperation among students, (3) uses active learning
techniques, (4) gives prompt feedback, (5) emphasizes time on task, (6) communicates high expectations, and (7) respects diverse talents and ways of learning. Frequently used

Table 1

*Characteristics of an Excellent Teacher*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content expertise</td>
<td>Obviously a faculty member must be knowledgeable in the content field in order to teach it. However, excellent content expertise in a <em>necessary</em> but <em>insufficient</em> quality for teaching excellence.</td>
</tr>
<tr>
<td>Affective traits/skills</td>
<td>▪ Enjoy teaching as much or more than they enjoy working in their field.</td>
</tr>
<tr>
<td></td>
<td>▪ Model the best characteristics of an accomplished <em>practitioner</em> in the fields they are teaching.</td>
</tr>
<tr>
<td></td>
<td>▪ Model the best characteristics of a life-long learner.</td>
</tr>
<tr>
<td></td>
<td>▪ Is demanding but fair.</td>
</tr>
<tr>
<td></td>
<td>▪ Is ethical and honest.</td>
</tr>
<tr>
<td></td>
<td>▪ Is comfortable admitting ignorance.</td>
</tr>
<tr>
<td>Performance skills</td>
<td>▪ Speaks clearly.</td>
</tr>
<tr>
<td></td>
<td>▪ Is organized when making a presentation.</td>
</tr>
<tr>
<td></td>
<td>▪ Uses examples when teaching.</td>
</tr>
<tr>
<td></td>
<td>▪ Uses humor effectively.</td>
</tr>
<tr>
<td></td>
<td>▪ Creates an appropriate psychological environment for learning.</td>
</tr>
<tr>
<td>Cognitive skills</td>
<td>▪ <em>Instructional design</em>—develops and uses learning objectives in designing effective learning experiences;</td>
</tr>
<tr>
<td></td>
<td>▪ <em>Instructional delivery</em>—skilled in presenting information in a variety of delivery modes</td>
</tr>
<tr>
<td></td>
<td>▪ <em>Instructional assessment</em>—skilled in the design and use of a variety of tools and procedures to assess student learning.</td>
</tr>
</tbody>
</table>
by scholars of postsecondary teaching and learning to assess teaching practices and pedagogical innovations (Nufer, 2003, for instance), these seven principles might also serve as the basis of a content review. Ultimately, in the absence of a widely agreed upon definition of effective college teaching, institutions that seek to validate their student rating instruments would need to articulate their own depiction of the trait(s) intended to be measured by each of the items in their chosen instrument.

Another validation activity that can provide evidence based on test content is examining the number and kinds of items included in an instrument, as the “items [of a multidimensional ratings form] must elicit a representative sample of student descriptions from the relevant domains of instructor behavior” (Abrami, 1989, p. 223). As Abrami (1989) puts it, “A rating form should not contain too many items assessing onesort of instructor behavior and too few items assessing another; …[it] should contain items equally relevant to each of the instructional situations for which it was designed” (p. 223). The issue of which behaviors are measured is particularly important when the resultant data will be used to provide teachers with formative feedback about their teaching (Marsh & Roche, 1993). Moreover, as Franklin (2005) reminds us, the majority of the ratings instruments in use were developed when lectures were the primary instructional method, so that today’s distance learning courses and active pedagogies may necessitate reevaluations of survey contents.

**Evidence based on internal structure.** Previously considered part of “construct validity,” this validation approach examines the statistical properties of the items and the extent to which these properties reflect the intended properties of the instrument and items. Some studies of internal structure have attempted to identify a “common” set of
factors that underlie the construct under examination: instructional quality (Ory & Ryan, 2001). For instance, Feldman’s (1976) meta-analysis of 23 studies resulted in a list of twenty-two common dimensions of instruction found on dozens of rating instruments. A limitation in using this approach to validation, Ory and Ryan (2001) point out, is that the findings tell us about how students respond to particular items generally included on ratings forms, but drawing conclusions based on these studies is restricted by the absence of a clear understanding of what they represent. Item analysis procedures and differential item functioning (DIF) studies can also be used to obtain evidence based on internal structure, as they can provide indices of item difficulty, discrimination, and fit.

**Criterion-related evidence.** This broad category of evidence covers several of the types of validity described in the previous validation framework: criterion validity and most of what was called construct validity, including convergent and discriminant validity (Goodwin & Leech, 2003). In Ory and Ryan’s (2001) words, criterion-related evidence reflects the questions “Is there a relationship between student ratings and variables external to the rating forms that are (1) expected to be predicted by the ratings and (2) similar measures of the intended construct” (p. 35). Studies examining the convergent and discriminant validity of student ratings aim to demonstrate that they “correlate relatively well with measures based on other methods for assessing the construct of quality of instruction and relatively less well with measures assumed to assess other constructs” (Greenwald, 1997, p. 1184). To do so, researchers have used multi-trait multi-method studies which correlate student ratings with select other measures of effective instruction across a variety of courses (Ory, 2001). They have found high positive correlations between student ratings and alumni ratings (Centra,
1974; Overall and Marsh, 1980), observer ratings (Feldman, 1989; Murray, 1983), and student comments (Ory, Braskamp, & Pieper, 1980), as well as moderate correlations between student ratings and peer ratings (Feldman, 1989).

With respect to discriminant validity, one of the most contentious relationships examined to date is that between student ratings and students’ (actual or expected) grades, leading to the pervasive myth that instructors can easily “buy” good ratings by awarding high grades. And it is the case that students who expect high grades rate instructors more favorably than students who expect low grades (Abrami, Dickens, Perry, & Leventhal, 1980; Feldman, 1976; Howard & Maxwell, 1980; Marsh, 1987), ostensibly validating the myth. At the University of Illinois at Urbana-Champaign, Ory (2001) found a correlation of approximately .30 between expected grades and overall instructor and course ratings. Then again, using two alternative models to describe the relationship between grades and student satisfaction, Howard and Maxwell (1980) found that the direct causal influence of grades on satisfaction appears to be minimal. “Rather than simply as evidence of contamination due to grading leniency,” they argue, “the relationship between grades and student satisfaction might be viewed as a welcomed result of important causal relationships among other variables” related to student motivation (p. 819). In light of these conflicting perspectives, it may be useful to control for the variance in ratings attributable to grades whenever possible.

Discriminant validity studies also encompass the copious bias studies of student ratings that have been conducted, those so intent on identifying extraneous influences on survey data that March (1987) and others have referred to them as the “witch hunt for potential biases in students’ evaluations” (p. 305). Tied with grades in the mythology of
student ratings is the notion that ratings data reflect an instructor’s personality, that ratings are a popularity contest of sorts. Although most teaching situations entail an element of performance (Lowman, 1995), scholars have determined that an instructor’s innate personality does not significantly influence student ratings. As Theall and Franklin (2001) state plainly, “There is no basis for this argument and no research to support it” (p. 49). Instead, research suggests that instructors’ actions in the classroom bear considerably more on ratings than any of their personal traits (Erdle, Murray, & Rushton, 1985).

Other instructor characteristics that have been examined and found not to be significantly correlated to student ratings, include gender, rank, age, years of experience, and research productivity (synthesized in Aleamoni, 1999 and Ory, 2001). In contrast, professors tend to secure higher ratings than teaching assistants; ratings in elective courses tend to be higher than those in required courses; ratings in higher-level courses tend to be higher than in lower-level courses; and ratings appear to differ consistently (but not substantially) based on disciplinary area (Ory, 2001). “In descending order, the disciplines are arts and humanities, biological and social sciences, business, computer science, math, engineering, and physical science” (Ory, 2001, p. 5). Evidence on the impact of class size is mixed (Alemoni, 1999). After reviewing 73 studies published prior to 1987 and 81 additional ones published between 1987 and 1999, Aleamoni (1999) concluded that all of this research supports the fact that “student rating myths are (on the whole) myths” (p. 159). It may be time to call off the witch hunt.

Multi-section studies have examined the relationship between student ratings and the most frequently researched criterion variable linked to these ratings: student achievement. Examining the final exam scores of students enrolled in multiple sections of
the same course taught by different instructors, these studies reflect the fact that (as mentioned) student learning is considered the most widely accepted criterion of effective teaching. They attempt to respond to the important, long-standing question, “Do higher ratings indicate that students learned more?” Scholars have, in fact, identified substantial correlations between student ratings and their achievement. Cohen’s (1981) oft-cited meta-analysis of all known multi-section validity studies found correlations between student achievement and student ratings of skill, overall course, structure, student progress, and overall instructor. Yet Feldman (1997) points out that early multi-section studies tended to measure lower-level educational objectives such as memory, as opposed to the higher-order outcomes more emphasized in academia today, and several others have pointed out that good examination scores can often be the result of unethical or poor teaching practices, such as teaching to the test (Carrell & West, 2010; Scriven, 1983).

Evidence based on relations to other variables also includes studies of validity generalization. “Would the relationships, or lack thereof, found between student ratings and various external variables exist in different settings with different students at different times?” (Ory & Ryan, 2001). In response, Abrami, d’Apollonia, and Cohen (1990) recognized that many of the multi-section studies were conducted in introductory courses, such that the correlations between ratings data and student achievement may not be as strong in upper-level courses where teaching behaviors may differ considerably. Research provides more definitive evidence of generalizability in relation to biasing influences since these types of studies were conducted in many different institutions, with different types of students, courses, and content (Ory & Ryan, 2001).
Evidence based on response processes. This category of validation evidence examines the relationship between the experience of responding to an instrument and the construct one intends to measure. In other words, whether the nature of the student rating process (i.e. filling out an online or in-person survey at the end of a semester consisting of x number of Likert-scale type questions, etc.) corresponds well to the measurement of teacher quality. It would be relevant to know, for instance, what motivates students to complete these surveys, who they think will have access to the survey data, whether/how they think their feedback will be utilized, whether they are implicitly comparing one instructor to another when responding. Moreover, one would need to explore how students interpret the rating scale being used for evidence based on response processes. Some students may not select a perfect “5” or “strongly agree” because they think these represent a level of perfection that is simply unattainable (Ory & Ryan, 2001). As Ory and Ryan (2001) concluded, very little is known about the nature of the students’ response processes as they complete rating forms, but learning more about this process would improve our analysis and use of student ratings.

Goodwin and Leech (2003) indicate that “also included [in evidence based on response processes] are investigations of the ways in which observers, judges, and raters use criteria to record and evaluate behavior, performances, essays, and so forth” (p. 184). Applied to student ratings, this point is illustrated by Angelo and Cross’s (1993) suggestion that when someone asks, “Are student ratings valid?” they really mean, “Are students really good judges of effective teaching?” The myth is that students are not qualified to rate instructors, but individuals more familiar with the research and practice of evaluation point out that students are well qualified in many instances (Theall &
Sciven (2005), who has written at length about potential sources of validity for student ratings, explains that students are in a unique position to rate their instructors: “Students are the most frequent observers of many facets of teaching and learning and their collective opinions as witnesses can provide useful information, particularly when they are asked to observe specific behaviors or materials” (p. 16).

Weimer (2010) concurs:

As important as the view from the front of the room is, the experience in the desk is just as important, and the front of the room looks different when viewed from the desk. Students are in a good position to offer feedback. They are there for the course from start to finish; their experience is first-hand and fresh. They can say better than anyone else whether the course design and teacher actions motivated and expedited their learning. (p. 51)

Scriven adds that the student views provided by ratings are also relevant because they reflect tastes and preferences about teaching and learning we would not otherwise be able to access, and as any retention scholar would underscore, understanding students’ tastes and preferences is crucial to promoting their persistence.

**Evidence based on the consequences of testing.** The final category of evidence involves the consequences of using an instrument, which is particularly significant when the resultant data will be used in high-stakes decision-making. As pertains to student ratings, Ory and Ryan (2001) express the validation question concerning consequences as “How does evidence of the intended and unintended consequences of testing inform validity decisions and our use of the tests?” (p. 37). To this end, they provide a thorough inventory of potential and actual, intended and unintended consequences of student ratings (replicated in Appendix A). The intended consequences reflect the primary uses of ratings data outlined above (to help students select courses, to make promotion/tenure
decisions, to provide feedback about teaching). Unintended consequences listed by Ory and Ryan (2001) include “teaching to the ratings”—that is, instructors adjust their teaching (by assigning less reading, for instance) in order to improve ratings; ratings are used as the only measure of faculty effectiveness; and students reward poor teaching in return for higher grades. Each institution would thus be wise to clearly articulate the intended consequences of its student rating practices and attempt to anticipate and preempt unintended consequences. Its ratings data will be valid only to the extent that the intended consequences are aligned with actual practice.

In addition, it is also noteworthy that most objections to ratings data have dealt with their unintended use—or misuse—as the basis for high-stakes administrative decisions such as promotion and tenure. Cohen (1980) affirmed, however, that “there is little controversy over their use for purposes of improving instruction” (p. 322), the primary function which motivated the current study. In fact, Murray (1987) provides four reasons why ratings can improve teaching effectiveness: (a) They provide useful feedback for diagnosing an instructor’s strengths and weaknesses; (b) feedback can provide the impetus for a faculty member’s professional development; (c) their use in personnel decisions offers a tangible incentive to enhance one’s teaching; and (d) their use in tenure decisions means good teachers more likely to be retained. Looking through the lens of the contemporary validity framework reveals that “in content, [response process], and consequential validity…little evidence has been provided or even looked for” (Ory & Ryan, 2001, p. 40), but also that a substantial amount of evidence has been amassed with regard to the internal structure of student ratings as well as their
relationship to other variables. All of this suggests that ratings data provide valuable insight about instructor efficacy that, used effectively, can promote enhanced teaching.

**Rationale and Research Questions**

Before describing the current study in more detail, I will turn again momentarily to empirical research on student retention to describe a study similar to mine. Noting that “no previous research [had] yet examined the impact of student course ratings” on student retention, Langbein and Snider (1999) hypothesized that, “at least in part, student ratings are a measure of customer satisfaction” such that “course ratings should have an independent impact on the decision to continue to purchase the educational good and services offered by the students’ institution, or to exit the institution” (p. 459). Using logistic regression to examine data from a private, four-year liberal arts university in the Northeast, they found support for their hypothesis that enrollment in consistently poorly rated courses significantly reduces the probability of retention. However, they also reported the seemingly-counterintuitive finding that students enrolled in the best-rated courses appear just as likely as those in the lowest-rated courses to leave the institution, suggesting to Langbein and Snider that there is an optimal range of student rating scores in terms of maximizing the probability of retention.

In much the same way, I explored the relationship between students’ average course ratings and student retention at the University of Miami. My specific research questions were thus:

Q1. To what extent do class average GLOBAL ratings of courses taken by a given student predict his/her likelihood of graduating within six years?
Q2. To what extent does the effect of class average GLOBAL ratings on graduation within six years depend on students’ race/ethnicity and/or total grade-point average?

Q3. To what extent do class average ratings of instructor pedagogical behaviors/skills (presentation, generates interest, availability) for courses taken by a given student predict his/her likelihood of graduating within six years?

Q4. To what extent does the effect of class average ratings of instructor pedagogical behaviors/skills (presentation, generates interest, availability) on graduation within six years depend on students’ race/ethnicity and/or total grade-point average?

GLOBAL refers to two items included in the University of Miami’s end-of-course student survey form which provide a student’s overall estimations of the instructor: (a) “My overall evaluation of the instructor is positive;” and (b) “I would recommend this instructor to a friend.” Several ratings scholars have suggested that general questions about the overall effectiveness of the course and instructor are good predictors of the overall quality of the course/instructor (Cashin & Downey, 1992; d’Apollonia & Abrami, 1997).

The particular pedagogical behaviors examined are those embedded in the three additional questions posed to students in UM’s core survey items: (a) “The instructor presents course material effectively;” (b) “The instructor stimulates interest in the course;” (c) “The instructor is available and willing to meet with students.” These items resonate well with the conceptions of effective college teaching provided by both Arreola
(2005), and Chickering and Gamson (1987) described above. This study was also limited
to these ratings items, namely those in UM’s core set of questions, because they are the
ones for which the greatest amount of data were available, facilitating a comparison
amongst a great number and variety of students. Why were interactions based on
race/ethnicity and total GPA selected for inclusion? Race/ethnicity was included because
the completion rates of students from underrepresented minority groups students are so
troubling, and because it was conceptually plausible that the impact of ratings on
retention could vary based on this trait. Total GPA was included because of the long-
established association between students’ grades and their instructor ratings, as well as
between students’ grades and persistence—both of which endorse the possibility that the
impact of ratings on retention could vary based on students’ total grade-point averages.

In Langbein and Snider’s (1999) study, retention was defined as “whether a
student who entered as a first semester freshman or as a freshman/sophomore transfer
graduated or returned to [the college] four semesters later” (p. 460). Although this
methodological decision does take into account the importance of first to second year
persistence, it is limited in that it provides no information about students’ degree
attainment. The use of six-year graduation as the primary outcome variable in the study
better reflects the fact that retention is a key University of Miami objective and
graduation the goal of most of its undergraduates. Moreover, Langbein and Snider’s
(1999) use of a composite course rating average may have obscured the impact of
specific instructor behaviors, an impact which would be extremely useful in determining
how to improve an institution’s teaching practices.
One limitation of both Langbein and Snider (1999)’s research and my own is that neither one examined an individual student’s ratings of his/her instructors to determine whether these appraisals predict his/her retention. This is because student ratings at most institutions are anonymous. I instead utilized (just as Langbein and Snider did) the mean ratings of the courses in which an individual student was enrolled, with the addition of mean ratings for specific instructor pedagogical behaviors/skills. As currently configured, then, student ratings were treated as an important contextual variable reflecting students’ classroom experiences and interactions with faculty members, and the study aimed to determine the impact of this contextual, group-level variable on an individual-level outcome—that is, students’ ability to graduate within six years.

This study also extends the aforementioned work by Braxton, Bray, and Berger (2000) and Pascarella, Seifert, and Whitt (2008) on the relationship between students’ perceptions of their instructors and their degree completion. As alluded to previously, a key distinction between their work and my investigation is that I used data from actual student ratings, data which are already collected on most U.S. college campuses, as opposed to conducting an additional survey. This cost-effective approach allows an institution to track changes in student ratings over time and to integrate these data into existing enrollment management models. In addition, using actual student ratings allowed me to explore the impact of several instructor behaviors as well as a global assessment, not simply the two measures (organization/preparation and instructional skill/clarity) included in Braxton, Bray, and Berger (2000) and Pascarella, Seifert, and Whitt (2008)’s studies—which, in turn, begot specific implications for improving student retention at the University of Miami, described in Chapter Five.
Chapter Three: Method

This chapter first presents the setting of the study and the sample, including how the latter was selected. It proceeds to describe each of the measures utilized, providing a brief rationale for the inclusion of each, and then reports the descriptive statistics for the sample. The methodology used in model building is next described, and the statistical approach used to examine each of the four research questions is detailed.

The study at hand examined data from the University of Miami, a private research university with more than 15,000 students located in the southeast. The demographic, pre-entry, and University student data (including student ratings) came from a variety of sources, including the Registrar’s Office and the Office of Testing and Evaluation Services, and they were all compiled and made available by the University of Miami’s Office of Planning, Institutional Research, and Assessment. It is also necessary to note that this University has had an enrollment management philosophy and set of practices in place for more than 25 years (Orehovec & Associates, 2011).  

Sample

The sample was selected based on enrollment at the institution in the years 2002, 2003, and 2004 so that six-year graduation rates could be examined. It therefore consisted of University of Miami undergraduate students who enrolled as first-semester freshmen in the fall of 2002, 2003, and 2004, and (to allow for comparison) for whom 12 or more of the courses they rated used the University’s core survey items.  

---

5 An explanation of enrollment management is provided on pages 22 to 26.
6 Although the University of Miami Testing Center encourages Schools and Colleges to use the UM core items, which were initiated in the spring of 1988, these entities can opt to use their own survey forms to meet their unique needs. Also, some courses simply are not rated, for various reasons. This cut-off figure of 12 was selected after reviewing the ratings data provided and in light of previous research on student ratings. Potential limitations ensuing from this decision are discussed in Chapter Five; see page 100.
Business uses a different set of core items, students enrolled in this school were excluded from the study. All in all, the working sample amounted to slightly upwards of four thousand (n=4,239). Of these, 59% of students were female students and 41% male. With respect to their racial/ethnic composition, 55% were classified as non-Hispanic white, 23% Hispanic, 12% non-Hispanic Black, and 7% Asian/Pacific Islander.

**Measures**

**Dependent variables.** The outcome variable of interest was graduation within six years. If the student entered as a first-semester freshman in 2002 and graduated in or before the spring 2008 semester, the student was considered “graduated” (coded 1); if the student did not graduate during this period, “not graduated” (coded 0). This was repeated for the two following student cohorts. The study did not differentiate among students who graduated within three, four, five, or six years. It was only concerned with whether or not students were able to attain their degree within six years, which is a commonly used benchmark institutionalized by the federal “Student-Right-to-Know and Campus Security Act” passed in 1990 (Selingo, 2012). As reported in Table 2 below, 78% of students in the sample graduated from the University within six years. This figure is considerably higher than the national averages for degree completion described above, yet it

---

Incidentally, the testing Center announced in the spring of 2011 that it was piloting a new general form consisting of eleven questions.

7 The University of Miami has 12 schools and colleges (Architecture, Arts & Sciences, Business, Communication, Continuing Studies, Education, Engineering, Law, Medicine, Music, Nursing, and Rosenstiel School of Marine & Atmospheric Science).

8 The proportion of missing cases for each independent variable and for the outcome variable changes the analytical sample, so it is necessary to acknowledge this factor. The number of missing cases ranged from 8 (in the case of “total GPA”) to 1135 (in the case of paternal level of education), and they are reported in Appendix C, Table C1. The working sample size of 4239 also reflects the decision to exclude cases in which students’ families had adjusted gross income values less than 0, as these values skewed the data considerably and made it difficult to interpret the results of the regression analyses.
approximates recent estimates of 80% six-year retention rates at private institutions (Chronicle Almanac, 2011).

**Independent variables.** For reasons explained in the preceding chapter, the main independent variables of interest in the study are student ratings. The student rating data utilized were collected between the years 2002-2010 through the University of Miami’s end-of-course faculty evaluation surveys which employ a Likert Scale of strongly agree to strongly disagree. The University of Miami’s Testing and Evaluation Services unit combines the percentage of students responding “strongly agree” and “agree,” coded as “yes,” and the percentage of “strongly disagree” and “disagree” responses, coded as “no.” The current study used only the affirmative responses—that is, the percentage of “yes’s”—in the analyses given that these were highly correlated with the “no’s.” In particular, students’ responses to these five core/common items were examined:

1. My overall evaluation of the instructor is positive.
2. I would recommend this instructor to a friend.
3. The instructor presents course material effectively.
4. The instructor stimulates interest in the course.
5. The instructor is available and willing to meet with students.

Since Q1 and Q2 above are both what are known as GLOBAL items, students’ responses to them were averaged, leaving a total of four items to be examined. The first item will be subsequently referred to as GLOBAL; the second will be called PRESENTATION; the third, INTEREST; and the fourth, AVAILABILITY.

---

9 This decision warrants discussion in the limitations section of Chapter Five; see page 100.
10 Items 1 and 2 had a correlation of .68, p < .01.
To obtain the rating scores, means were computed for the percentage of affirmative responses for each of the core items, for all of the classes taken by each student during their $m$ number of years as University of Miami undergraduates. The descriptive statistics for the sample (provided in Table 2 below) indicated that mean % affirmative instructor ratings were favorable. Ranging from 79 to 82%, the class averages of students’ overall impression of their instructors and estimation of the three given pedagogical behaviors/skills were, on the whole, positive. Means of GLOBAL exhibited the highest standard deviation (8.09%), and INTEREST had the highest range (78.5%).

**Covariates.** Covariates consisted of demographic, pre-college, and other factors expected to influence student retention. Although the literature is inconclusive with respect to the influence of gender (Astin, 1975; Reason 2001; Reason, 2003; Tinto,

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLOBAL</td>
<td>82.09</td>
<td>83.10</td>
<td>6.43</td>
<td>78.50</td>
<td>-1.84 (SD=.04)</td>
<td>8.29 (SD=.08)</td>
</tr>
<tr>
<td>PRESENTATION</td>
<td>81.00</td>
<td>81.12</td>
<td>6.09</td>
<td>74.00</td>
<td>-0.62 (SD=.04)</td>
<td>4.48 (SD=.08)</td>
</tr>
<tr>
<td>INTEREST</td>
<td>78.75</td>
<td>79.64</td>
<td>8.09</td>
<td>70.00</td>
<td>-0.66 (SD=.04)</td>
<td>1.17 (SD=.08)</td>
</tr>
<tr>
<td>AVAILABILITY</td>
<td>81.00</td>
<td>81.39</td>
<td>5.52</td>
<td>63.00</td>
<td>-0.71 (SD=.04)</td>
<td>3.43 (SD=.08)</td>
</tr>
</tbody>
</table>

*Note. GLOBAL=mean course ratings of overall instructor; PRESENTATION= mean course ratings of instructors’ presentation skills; INTEREST = mean course ratings of instructors’ ability to generate interest in course; and AVAILABILITY = mean course ratings for extent to which instructors were available outside of class.*
1987), this variable was nonetheless included as a key sociodemographic trait. Males were coded 1 and females 2.\textsuperscript{11} Students’ race/ethnicity was incorporated given that strong correlations have been found between this sociodemographic feature and student persistence, with White and Asian students persisting at considerably higher rates than Hispanic and Black students (Astin, 1997; Braxton, Brier, & Hossler, 1988; Murtaugh et al., 1999). The University of Miami’s high enrollment of Hispanic students allowed for a comparison between the experiences of Hispanic and non-Hispanic White students. In the University of Miami’s student records, this variable consists of six levels (non-Hispanic White, Hispanic, non-Hispanic Black, Asian/Pacific Islander, American Indian, and unknown), so it was dummy coded, using White as the reference group. Given that there were too few students identified as American Indian in the sample, this category was combined with the unknown category.\textsuperscript{12}

High-school grade-point average, considered the strongest precollege predictor of college retention (ACT, 2007), and family adjusted gross income, considered the second highest predictor of college retention (ACT 2004), were controlled for in an attempt to isolate the unique contribution of student ratings to variations in student persistence. SAT scores, another precollege variable, are also considered a significant predictor of retention (Astin, Korn, & Green, 1987; Tross Harper, Osher, & Kneidinger, 2000), so they were likewise controlled for in an attempt to isolate the unique contribution of student ratings to variations in student persistence. High school GPA, income, and SAT scores are all

\textsuperscript{11} A summary of variable definitions and coding is provided in Table B1 (in Appendix B).
\textsuperscript{12} The decision to combine the American Indian and unknown categories was one borne from necessity, in order to minimize the number of categories in this variable. Since there was no conceptual basis for this grouping and the category has no conceptual value, all discussion of results pertaining to students’ race/ethnicity will focus on the other racial and ethnic categories.
continuous variables. Given that first-generation college students have consistently experienced higher rates of departure (Ishitani, 2006; Nuñez & Cuccaro-Alamin, 1998), it was important to include parental years of schooling as a variable. This variable, originally containing six categories, was recoded into four categories by combining the non-high school graduates and the high school graduates, as well as the “some college” and the college degree completion categories. The resulting four categories were dummy coded, using “non-high school graduates/high school graduates” as the reference category. This classification scheme was intended to allow for comparison with previous research findings on first-generation students (Inkelas, Daver, Vogt, & Leonard, 2007).

One institution-related variable whose influence on graduation was examined represents the financial aid provided to each student, in light of research which suggests that some departure decisions may be linked to students’ finances (DesJardins et al., 2002, Knight & Arnold, 2000). Two variables were used as markers of financial aid: (a) FA$_a$, which stands for the ratio of unmet need to gross need, and (b) FA$_b$, which stands for the ratio of total scholarships and grants to gross need. College grade-point average, the most powerful predictor of retention, signals students’ level of academic integration and is known to be correlated with instructor ratings (Murtaugh, Burns, & Schuster, 1999; Pascarella & Chapman, 1983; Pascarella & Terenzini, 1978). This continuous variable was a notable one in the study because it was the only in-college academic variable with which to compare the impact of the student ratings variables.

---

13 The ratio of unmet need to gross need is indicative of how much of a student’s financial need (as determined by the federal government, through the Free Application for Federal Student Aid—the FAFSA) has been met by the institution in the financial aid package designed for that student.
Table 3

*Descriptive Statistics for Categorical Variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduation</td>
<td>Graduated within 6 yrs.</td>
<td>3313</td>
<td>78.2</td>
</tr>
<tr>
<td></td>
<td>Did not</td>
<td>926</td>
<td>21.8</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>1741</td>
<td>41.1</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2498</td>
<td>58.9</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Non-Hispanic White</td>
<td>2327</td>
<td>54.9</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>966</td>
<td>22.8</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic Black</td>
<td>502</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td>Asian/Pacific Islander + other/unknown</td>
<td>444</td>
<td>10.5</td>
</tr>
<tr>
<td>Father’s educational level</td>
<td>Non h.s.grads &amp; h.s.grads</td>
<td>541</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>Some college &amp; BA degree</td>
<td>1593</td>
<td>37.6</td>
</tr>
<tr>
<td></td>
<td>Post-BA</td>
<td>970</td>
<td>22.9</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>1135</td>
<td>26.8</td>
</tr>
<tr>
<td>Mother’s educational level</td>
<td>Non h.s.grads &amp; h.s.grads</td>
<td>599</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>Some college &amp; BA degree</td>
<td>1990</td>
<td>46.9</td>
</tr>
<tr>
<td></td>
<td>Post-BA</td>
<td>716</td>
<td>16.9</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>934</td>
<td>22.0</td>
</tr>
</tbody>
</table>

Descriptive statistics for the sample, shown in Tables 4 and 5, indicated that students had a mean high-school grade-point average of 4.11, mean SAT scores of 1235, and household incomes averaging $92,000. About five eighths of them were female, and
nearly two-thirds of their parents attended college. Most measures were normally
distributed. The most notable exceptions is the second financial aid variable (FAb—the

Table 4

Descriptive Statistics for Continuous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.S. GPA</td>
<td>4.11</td>
<td>4.05</td>
<td>0.56</td>
<td>3.86</td>
<td>0.23</td>
<td>0.48</td>
</tr>
<tr>
<td>SAT</td>
<td>1235</td>
<td>1230</td>
<td>129.57</td>
<td>820</td>
<td>-0.08</td>
<td>-0.14</td>
</tr>
<tr>
<td>FAa</td>
<td>.18</td>
<td>.11</td>
<td>.21</td>
<td>1.00</td>
<td>1.28</td>
<td>1.29</td>
</tr>
<tr>
<td>FAb</td>
<td>1.29</td>
<td>.60</td>
<td>5.58</td>
<td>179.44</td>
<td>22.72</td>
<td>624.93</td>
</tr>
<tr>
<td>Income</td>
<td>92,467</td>
<td>70,788</td>
<td>9.41</td>
<td>999999</td>
<td>4.05</td>
<td>27.35</td>
</tr>
<tr>
<td>1st semester GPA</td>
<td>3.23</td>
<td>3.36</td>
<td>0.63</td>
<td>4.0</td>
<td>-1.48</td>
<td>3.42</td>
</tr>
<tr>
<td>Total GPA</td>
<td>3.24</td>
<td>3.36</td>
<td>0.58</td>
<td>4.0</td>
<td>-1.60</td>
<td>4.07</td>
</tr>
</tbody>
</table>

Note: H.S. GPA=students’ high school grade-point averages; SAT= score on SAT exam; FAa =
ratio of unmet need to gross need; FAb = ratio of total scholarships and grants over gross need;
income = family adjusted gross income; 1st semester GPA = first semester grade-point average;
and total GPA= student’s grade-point average over m years at UM.

ratio of total scholarships and grants to gross need), which was positively skewed (with a
kurtosis of 625).14 Six variables were skewed more than 1.0.

Analyses

The model building stage, which was informed by both the extensive
research/theoretical considerations described above as well as by empirical factors,

14This would be problematic and could compromise the reliability of the results in linear regression
analysis, but this is not the case with logistic regression, which can accommodate skewed variables. In the
case of FAb, the high kurtosis reflects the preponderance of zero values which result in a notable peak.
proceeded with a series of steps. As Hosmer and Lemeshow (2000) explain, it is best to minimize the number of variables in a model so that “the resultant model is more likely to be numerically stable, and is more easily generalized” (p. 92). To develop a parsimonious model, then, I heeded their suggestion and began the variable selection process with “a careful univariable analysis of each variable” (p. 92). Since the independent variables in this study are both categorical and continuous, this step required two different approaches: (a) A chi-square test of independence, complemented by Cramer’s $V$, was used to determine whether—and the extent to which—graduation varied with each categorical variable; and (b) a $t$-test, complemented by Cohen’s $d$, was utilized to examine the relationship between each of the continuous independent variables and graduation.

The information derived from these analyses was instrumental in determining which variables to include in the final model. However, it was necessary to examine how each of the independent variables relates to the other independent variables, so as to identify highly correlated predictors and prevent multicollinearity. Again, the chi-square test of independence, complemented by Cramer’s $V$, was used to examine the relationship between pairs of categorical variables (for instance, academic discipline and race/ethnicity). Both $t$-tests, complemented by Cohen’s $d$, and ANOVA were utilized to examine the relationship between categorical/dichotomous variables and continuous variables. Pearson’s bivariate correlation was used to examine the relationship between two continuous variables. In the event that two variables were highly related, I determined which one to include in the model based on both theory and the results of the chi-square tests, $t$-tests, AOVAs, and correlations.
The next stage in model building proposed by Hosmer and Lemeshow (2000) involves a multivariate analysis. This entailed using logistic regression with all of the variables I decided to include based on the previous steps and the literature on student retention, and it also required examining possible interactions among the variables, as warranted by the literature base. Those variables found to be significant predictors of the outcome variable(s) and whose inclusion was supported by theory were incorporated in the final model.

All of these procedures facilitated an examination of my research questions. To explore research question 1, which asks, “To what extent do class average GLOBAL ratings of courses taken by a given student predict his/her likelihood of graduating within six years?” graduation was regressed on the mean GLOBAL rating score (across m years at UM) and all of the variables in the final model using logistic regression. Independent variables were entered in two blocks: Block one consisted of race/ethnicity, high school GPA, income, SAT scores, FAa, and total GPA, and block 2 consisted of GLOBAL.

To explore research question 2, “To what extent does the effect of class average ‘global’ ratings on graduation within six years depend on students’ race/ethnicity and/or total grade-point average?” graduation was regressed on the mean GLOBAL ratings, all of the variables in the final model, and each of the associated interaction terms. A combination of block and stepwise entry was utilized, and the order of block entry was as follows: Block one consisted of race/ethnicity, high school GPA, income, SAT scores, FAa, and total GPA; block 2 consisted of “global;” and block 3 consisted of the variables representing the interactions between GLOBAL and race/ethnicity, and between GLOBAL and total GPA.
To explore research question 3, “To what extent do class average ratings of instructor pedagogical behaviors/skills…for courses taken by a given student predict his/her likelihood of graduating within six years?” it was necessary to look closely at the results of the univariate analyses and the relationship among the three behavior variables themselves. This information determined that the behaviors (i.e., INTEREST, AVAILABILITY, and PRESENTATION) should be examined individually, and as a result, graduation was regressed on the mean rating score for each, one at a time, and all of the variables in the final model using logistic regression. Independent variables were entered in two blocks: Block one consisted of race/ethnicity, high school GPA, income, SAT scores, FAs, and total GPA, and block 2 consisted of the individual rating item—such that three regression models were examined.

Finally, to explore research question 4, “To what extent does the effect of class average ratings of instructor pedagogical behaviors/skills…on graduation within six years depend on students’ race/ethnicity and/or total grade-point average” graduation was regressed on the mean scores for each ratings variable, all of the variables in the final model, and each interaction term, one at a time. In this case, a combination of block and stepwise entry was utilized, and the order of block entry was as follows: Block 1 consisted of race/ethnicity, high school GPA, income, SAT scores, FAs, and total GPA; block 2 consisted of the individual pedagogical behavior being examined; and block 3 consisted of all of the interaction terms associated with the ratings variable entered in a stepwise process. Exploring research question 4 therefore required three regression analyses, one for each behavior rating, the results of which are presented in Chapter 4.
Assumptions of Logistic Regression

At this point it is worth noting that logistic regression has several assumptions, all of which were tested. The first of these is collinearity among predictors, that is, the independent variables cannot be linear combinations of each other. Checking the tolerance values and variance inflation factors (VIF) for each predictor indicated that they are within the acceptable ranges. That is, for VIF the threshold is commonly treated as ten, and all values were well below 2. Another assumption of logistic regression is that each observation is independent and, as such, that the error terms are independent. Using the Durbin-Watson statistic test for correlated residuals demonstrated that the statistic was close to two (1.917), which suggests that autocorrelation is likely not present.

Finally, logistic regression assumes that there is a linear relationship between continuous predictors and the logit of the outcome variable. One approach used to test this assumption is to create an interaction term for each continuous variable with its log transformations, enter these interaction terms into the regression model as additional predictors, and determine whether they are significant (Hosmer & Lemeshow, 1989). In the current study, this analysis determined that two of the interactions were significant, thereby violating the assumption of linearity for the two corresponding variables: total GPA and GLOBAL. Although this may have resulted in an underestimation of the degree of association between these predictors and graduation (i.e., more Type II errors), these predictors were nonetheless included, as logistic regression is robust to violations (Field, 2009).
Model Fit and Effect Size

Several approaches can be used to assess model fit when using logistic regression; however, the main statistic on which model fit analyses depend is the -2loglikelihood (-2LL). This is a “test of the significance of the difference between the likelihood ratio (-2LL) for the researcher’s model with predictors…minus the likelihood ratio for baseline model with only a constant in it” (Burns & Burns, 2008, p. 574-5). It therefore “measures the improvement in fit that the explanatory variables make compared to the null model” (Burns & Burns, 2008, p. 575). In the current study, the significance of the -2LL as determined by the model chi square was used as one indicator of model fit, and a reduction in the -2LL value upon addition of a variable was taken to mean that the variable improved the model’s ability to predict graduation. This was complemented by the Hosmer-Lemeshow goodness-of-fit statistic (HL). HL tests the null hypothesis that the model fits the data, such that a significant HL indicates that a model fails to predict the outcome with sufficient accuracy. The $R^2$ statistic, widely utilized in linear regression as a measure of goodness-of-fit as well as of how much variance in the outcome is accounted for by the model, has no analogous statistic in logistic regression (Sweet & Grace-Martin, 2002). However, pseudo $R^2$ measures have been developed, and one of these—Nagelkerke’s $R^2$—was used in this study as an additional gauge of model fit.

After considering model fit in logistic regression, interpretation consists of two main steps. Hosmer and Lemeshow (2000) explain that this involves “determining the functional relationship between the dependent variable and the independent variable, and appropriately defining the unit of change for the independent variable” (p. 47). Doing the former requires looking at the individual coefficients. However, these coefficients
represent the change in the logit that corresponds to a one-unit change in the predictor, making them difficult to interpret (Hosmer & Lemeshow, 2000). It is therefore advisable and customary to use the coefficients to identify the direction of the association (i.e., whether an increase in the predictor enhances or diminishes the odds of the outcome), and turn to the odds ratio (OR)\textsuperscript{15} to interpret the strength or meaningfulness of the association (Hosmer & Lemeshow, 2000). This approach is used in the chapter that follows to interpret the regression analyses results and thereby respond to the research questions at hand.

\textsuperscript{15} The OR—or exp(b)—represents the ratio of the odds of an event occurring in one group to the odds of it occurring in another group, and odds ratios are multiplicative, such that values greater than one indicate positive effects, and values between zero and one indicate negative effects.
Chapter Four: Results

I begin this chapter by presenting the results of the preliminary analyses, that is, those conducted before the regression analyses to determine which variables to include in the final regression models used to estimate the research questions. I then describe how well the baseline model predicted graduation and what the results of the baseline logistic regression analysis suggest about each variable’s impact on graduation. This is followed by the results of the univariate and logistic analyses utilized to investigate each of the research questions.

Preliminary Analyses

Preliminary analyses began by appraising the association between each independent variable and graduation. These analyses discerned statistically significant\(^{16}\) differences in graduation based on students’ race/ethnicity as well as their fathers’ educational levels. In the case of the race/ethnicity variable, \(V = .05, \chi^2(4) = 10.18, p = .037\), and the Cramer’s \(V\) value of .05 suggests that the magnitude of the association between race/ethnicity and graduation was small. In much the same way, the chi-square test was significant (\(V = 0.06, \chi^2(2) = 11.27, p = .037\)) for the father’s level of education, yet the Cramer’s \(V\) value of .06 suggests that the magnitude of this relationship was small.

Moving on to the continuous independent variables, statistically significant differences in graduation were identified based on high-school GPA, SAT scores, household income, the first financial aid variable (FA\(_a\)), both college GPA variables, as well as three out of the four ratings variables (Results are summarized below in Table 5).

\(^{16}\)In this study, a Type I error rate of .05 was used for all analyses.
Although all of the GPA variables were moderately associated with graduation within 6 years,\textsuperscript{17} the Cohen’s $d$ effect size values indicate that total college GPA had the strongest relationship with graduation: The total GPAs of students who graduated were 0.4 points higher than the GPAs of those who did not. In addition, the SAT scores of student who graduated were roughly 100 points higher than those of students who did not, and students who graduated within 6 years came from families with household incomes roughly $12,000 greater than those of students who did not graduate.

Table 5

*Results of t-tests and Cohen’s d Values for Relationship between Continuous Independent Variables and Graduation within 6 Years*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$t$</th>
<th>$df$</th>
<th>$d$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.S. GPA</td>
<td>-5.89</td>
<td>3806</td>
<td>-0.23</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>SAT</td>
<td>-2.16</td>
<td>4226</td>
<td>-0.08</td>
<td>.031</td>
</tr>
<tr>
<td>Income</td>
<td>-3.71</td>
<td>1544</td>
<td>-0.14</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>$F_{A_a}$</td>
<td>5.19</td>
<td>3433</td>
<td>0.22</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>1st semester GPA</td>
<td>-9.96</td>
<td>1200</td>
<td>-0.40</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Total GPA</td>
<td>-13.42</td>
<td>1123</td>
<td>-0.56</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Global</td>
<td>-10.57</td>
<td>1141</td>
<td>-0.44</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Presentation</td>
<td>2.20</td>
<td>1140</td>
<td>0.10</td>
<td>.028</td>
</tr>
<tr>
<td>Interest</td>
<td>-2.64</td>
<td>1199</td>
<td>-0.11</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

*Note.* Only statistically significant results are presented.

\textsuperscript{17}The most common way to interpret effect sizes is that proposed by Cohen (1988), whereby an effect size of about 0.20 is considered “small;” 0.50 is considered “medium;” and roughly 0.80 is considered “large.” However, interpretations of effect size vary by discipline, and in this study, correlations greater than or equal to .2 were considered noteworthy.
As mentioned in Chapter 3, additional preliminary analyses were conducted to examine the relationships among the independent variables themselves to detect highly correlated covariates. The correlation matrix in Appendix D (Table D1) illustrates the relationships among each of the continuous independent variables. Of note is that those students with higher high-school GPAs and SAT scores attained higher in-college GPAs, and that the correlation between high-school GPA and first-semester college GPA \((r(3797) = .314, p < .01)\) is nearly indistinguishable from that between SAT score and first-semester college GPA \((r(4214) = .311, p < .01)\). In the next series of tests, which looked at relationships among the categorical independent variables, the relationship between the maternal and paternal levels of educational attainment was significant \((V = 0.34, \chi^2(4) = 671.93, p < .01)\), with a Cramer’s \(V\) that suggests a small-to-medium relationship. (The results of the other chi-square tests are summarized in Table D2, in Appendix D.)

When examining relationships based on gender, a statistically significant difference was identified in the mean SAT scores of males and females \((d = 0.24, t(3469) = 7.51, p < .01)\) as well as in their high-school GPAs \((d = -0.36, t(3806) = -10.56)\): Male students’ SAT scores were slightly (about 30 points) higher than those of their female counterparts, yet females had higher high-school GPAs by about 0.20. Students’ in-college grade-point averages also differed by gender: For 1st semester GPAs, \(d = -0.16, t(3355) = -5.15, p < .01\), and for total GPA, \(d = -0.24, t(3271) = -7.72, p < .01\). In both cases, females attained higher GPAs (by 0.10 and 0.14 points, respectively).

Investigating the relationship between continuous variables and categorical variables with more than two levels required the use of ANOVA. The analysis examining
high school GPA as a function of race/ethnicity detected statistically significant differences \((F(4, 3803) = 75.68, p < .01)\), albeit with a small effect size \((\eta^2 = 0.07)\). A Bonferroni post-hoc test revealed that Hispanic students in the sample had higher high school GPAs than their white and black counterparts, whereas Asians students had higher high school GPAs than Hispanic, white, and black students. Statistically significant differences were also identified in the mean SAT scores per ethnic/racial group \((F(4, 4223) = 174.35, p < .01)\), and the effect size was meaningful \((\eta^2 = 0.14)\). Specifically, Asian students had the highest scores \((M=1287)\), followed by white \((M=1269)\), Hispanic \((M=1199)\), and black \((M=1129)\) students.

Statistically significant differences in household income as a function of race/ethnicity were also revealed \((F(4, 3646) = 42.81, p < .01, \eta^2 = .04)\) In particular, whites had household incomes 30 thousand dollars greater than those of Asian students, 37 thousand dollars greater than those of Hispanic students, and 67 thousand dollars greater than those of black students. Nevertheless, analysis of the first financial aid variable \((FA_a)\) suggests that white and black students were offered the most financial aid relative to their need. First-semester GPAs also varied by race/ethnicity at statistically significant levels \((F(4, 4221) = 21.91, p < .01, \eta^2 = .02)\). The 1\textsuperscript{st} semester GPAs of black students were lower than those of students from all other groups: 0.3 point lower than those of white students, 0.2 points lower than those of Hispanic students, and 0.4 points lower than those of Asian students. The 1\textsuperscript{st} semester GPAs of Asian students were higher (by 0.1 points) than those of Hispanic students. Total GPAs followed a similar pattern. The ANOVA detected statistically significant differences among the racial/ethnic groups \((F(4, 4222) = 55.40, p < .01)\), and the Bonferroni post hoc test revealed that the total
GPAs of whites and Asians were nearly equal, while the total GPAs of Hispanic and black students were lower by 0.1 and 0.4 points respectively.

The ANOVA results comparing students whose parents have varying levels of education attainment confirm previous research findings that individuals’ incomes increase commensurate with increases in degree completion. These analyses likewise point to slightly higher SAT scores, 1st semester GPAs, and total GPAs for students whose mothers or fathers attained comparatively higher levels of education.

Ultimately, taking into account the results of all preliminary analyses described above, I decided that the baseline model would consist of the following independent variables: race/ethnicity, high school GPA, income, SAT scores, the first financial aid variable (FAa), and total GPA. High school GPA, income, SAT scores, FAa, and total GPA were included because of the noteworthy effect sizes detected in the t-tests conducted to relate each of these to graduation. Total GPA was included instead of 1st semester GPA because the two are highly correlated, and total GPA seemed more appropriate given that the mean course ratings data stem from several years’ worth of courses. Although the magnitude of the association between race/ethnicity and graduation was small, this variable was included because the preliminary analysis identified it as a statistically significant predictor of graduation, and because there are such problematic discrepancies in degree attainment based on race/ethnicity nationwide.

**Tests of Model Fit and Baseline Interpretations**

As described in the preceding chapter, the main statistic on which model fit analyses is evaluated in logistic regression is the -2LL. In the baseline model (prior to the inclusion of any ratings variables), the -2LL value was 2996.903, which was
Table 6

Summary of Baseline Logistic Regression Model to Predict College Graduation

<table>
<thead>
<tr>
<th>Parameter estimated in the model</th>
<th>b</th>
<th>SE(b)</th>
<th>Wald $\chi^2$</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Block 0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.28</td>
<td>0.04</td>
<td>1176.00**</td>
<td>3.58</td>
</tr>
<tr>
<td><strong>Block 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.194</td>
<td>0.124</td>
<td>2.464</td>
<td>1.215</td>
</tr>
<tr>
<td>Black</td>
<td>0.231</td>
<td>0.150</td>
<td>2.378</td>
<td>1.260</td>
</tr>
<tr>
<td>Asian</td>
<td>0.314</td>
<td>0.198</td>
<td>2.530</td>
<td>1.369</td>
</tr>
<tr>
<td>HS GPA</td>
<td>0.012</td>
<td>0.094</td>
<td>.017</td>
<td>1.012</td>
</tr>
<tr>
<td>SAT</td>
<td>-0.011</td>
<td>0.004</td>
<td>7.135**</td>
<td>0.989</td>
</tr>
<tr>
<td>FA&lt;sub&gt;a&lt;/sub&gt;</td>
<td>-0.556</td>
<td>0.237</td>
<td>5.483*</td>
<td>0.574</td>
</tr>
<tr>
<td>Income</td>
<td>0.033</td>
<td>0.012</td>
<td>7.923**</td>
<td>1.033</td>
</tr>
<tr>
<td>Total GPA</td>
<td>1.053</td>
<td>0.083</td>
<td>158.982**</td>
<td>2.865</td>
</tr>
</tbody>
</table>

-2LL statistic: 2296.903
Model $\chi^2$: 223.02**
HL statistic: $p = .015$
Nagelkerke Pseudo $R^2$: .108

*Note.* “White” is the reference category for race/ethnicity. The SAT variable is in units of 10, and income is in units of $10,000. *$p$ < .05, **$p$ < .01.

1453.507 points lower than that of the intercept-only model, and the chi square test was significant ($\chi^2(9) = 223.016, p < .01$). All of this indicates that the addition of the covariates improved the model’s ability to predict graduation, corroborated by the Nagelkerke $R^2$ value of .108.\(^{18}\) However, the Hosmer-Lemeshow (HL) goodness-of-fit

\(^{18}\) It is important to note again that the *Nagelkerke $R^2$* value cannot be interpreted in a straightforward manner as is possible in linear regression. It does not indicate the amount of variance accounted for by the predictors, although it offers a rough approximation of fit and can be used in conjunction with the other omnibus tests described here (Long & Freese, 2006).
statistic was also significant \((p = .015)\), indicating that the model failed to predict graduation with sufficient accuracy to reach the appropriate HL statistic.

For the baseline model (illustrated in Table 6), the coefficients for family income, SAT, \(F_{A,n}\), and total GPA were all statistically significant at \(p < 0.05\). More specifically, a one-unit (i.e., $10,000) increase in family income was associated with a 3.3% increase in the odds of graduating. A one-unit decrease in \(F_{A,n}\), the ratio of unmet need to gross need (which means that a greater proportion of a student’s need was met), was associated with a 42.6% increase in the odds of being retained. The coefficient for total GPA was likewise positive, and the impact of an improved total college GPA is remarkable: In block one, a one unit/one point increase in total GPA increases the odds of graduating within 6 years by 187%. The coefficient for SAT scores was negative, suggesting that lower SAT scores were associated with greater odds of graduating, and the odds ratio of .989 indicates that a 10-point increase in SAT score diminishes the odds that a student will graduate by only 1.1%.

**Research Questions 1 and 2: The Impact of GLOBAL**

The first research question sought to determine the extent to which students’ GLOBAL course ratings—that is, their overall evaluations of their instructors—predict their ability to graduate within 6 years. It was thus important to first determine how the GLOBAL variable related to the graduation outcome variable, as well as how it was associated with the other predictors. The \(t\)-test examining the relationship between GLOBAL and graduation was statistically significant \((t(1141) = -10.57, p < .001)\), and the effect size of -0.44 is indicative of a moderate relationship between GLOBAL and graduation. The GLOBAL course ratings of students who graduated within 6 years were
about 3 percentage points higher than those of students who did not graduate. GLOBAL ratings varied neither as a function of gender nor race/ethnicity, but they were significantly correlated with 1st semester GPA ($r(4215) = .150, p < .01$) as well as total GPA ($r(4216) = .206, p < .01$).

Table 7

*Summary of Logistic Regression Model to Predict Impact of GLOBAL Student Ratings on Graduation*

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>SE(b)</th>
<th>Wald $\chi^2$</th>
<th>Odds Ratio</th>
<th>-2LL</th>
<th>Model $\chi^2$</th>
<th>HL</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2996.90</td>
<td>223.02**</td>
<td>$p = .015$</td>
<td>.108</td>
</tr>
<tr>
<td>Global</td>
<td>0.056</td>
<td>0.007</td>
<td>60.93**</td>
<td>1.057</td>
<td>2934.73</td>
<td>285.19**</td>
<td>$p = .023$</td>
<td>.137</td>
</tr>
<tr>
<td>global×totalGPA</td>
<td>0.047</td>
<td>0.009</td>
<td>26.41**</td>
<td>1.048</td>
<td>2908.98</td>
<td>310.94**</td>
<td>$p &lt; .01$</td>
<td>.148</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01. Predictors are not shown. Refer to Table 6 for detailed baseline data.

The multivariate results offer more insight into the unique relationship between GLOBAL ratings and graduation, controlling for the influences of race/ethnicity, high school GPA, income, SAT scores, FAF, and total GPA (See Table 7). Upon addition of the GLOBAL ratings variable, the -2LL decreased from 2996.903 in block one to 2934.731 in block two; the omnibus test was significant ($p < .01$); and the Nagelkerke $R^2$ value was .137—all of which indicate that this predictor improved the model fit. Although the HL statistic was significant ($p = .023$), the GLOBAL variable was a statistically significant predictor of graduation ($p < .01$) after controlling for the effects of students’ race/ethnicity, high school GPA, income, SAT scores, FAF, and total GPA. In
terms of effect size, the odds ratio value of 1.057 indicates that per one-unit (one
percentage point) increase in students’ GLOBAL course ratings (their overall appraisals
of their instructors), a student’s odds of graduating improved by 5.7%.

The second research question concerned the extent to which the impact of
GLOBAL course ratings on graduation varies for different groups of students. Therefore,
the interactions between GLOBAL and students’ race/ethnicity, as well as that between
GLOBAL and students’ total grade-pointaverages, were examined by adding these
variables to the logistic regression estimates in block three. Results indicated that the only
interaction variable that was statistically significant was “global\textsuperscript{\texttimes}totalGPA”—that is, the
variable representing the interaction between students’ GLOBAL course ratings and their
total grade-point averages. As illustrated in Table 7, the addition of this variable resulted
in a change in the $-2\text{LL}$ from 2934.73 to 2908.98, suggesting that it improves the model’s
ability to predict graduation, and the omnibus test was significant ($p < .01$). The
Nagelkerke $R^2$ value was .148, although the HL statistic was significant. All of this
suggests that the relationship between GLOBAL ratings and graduation varies based on
students’ total GPAs. The positive coefficient indicates that the relationship between this
interaction and graduation is positive, and the odds ratio of 1.048 indicates that students’
odds of being retained increase by 4.8% for every one unit increase in the interaction
between total GPA and GLOBAL In other words, students with high GPAs especially
benefitted (in terms of their persistence) from being enrolled in courses with high
GLOBAL ratings.
Research Questions 3 and 4: The Impact of Pedagogical Behaviors

Research questions 3 and 4 concerned the three instructor pedagogical behaviors/skills included in UM’s survey core: INTEREST (instructor’s ability to generate interest), AVAILABILITY (instructors’ willingness and availability to meet with students), and PRESENTATION (instructors’ presentation skills). It was necessary to first look at the relationship among the ratings themselves and thereby determine whether to create separate models for each. Indeed, the correlations among the INTEREST, AVAILABILITY, and PRESENTATION ratings variables were all statistically significant as well as considerably high. Ranging from .50 to .71, these Pearson’s values indicate that there is a strong association among the ratings variables, but they are not suggestive of multicollinearity; that is, none of these independent ratings variables appears to be a perfect linear combination of another, so each of them was individually examined. The sections that follow correspond to the three pedagogical behaviors/skills being investigated. Each section begins with the results of the regression analysis used to determine the unique contribution of each behavior/skill to graduation (i.e., research question 3) and is followed by the results of the interaction analysis for the given behavior/skill (i.e., research question 4).

Interest. As with GLOBAL it was important to consider how the INTEREST variable related to student graduation, as well as how it was associated with the other predictors. The t-test examining the relationship between INTEREST and graduation was statistically significant ($t(1198)=-2.637, p < .01$), and the effect size of -0.11 is indicative of a small relationship between INTEREST and graduation. The mean INTEREST course ratings of students who graduated within 6 years were about 1 percentage point higher
than those of students who did not graduate. INTEREST ratings were significantly correlated with students’ high school GPAs ($r(3797) = -.098, p < .01$) and SAT scores ($r(4217) = -.124, p < .01$). Although they did not vary based on students’ gender, INTEREST varied as a function of race/ethnicity ($F(4) = 1.70, p < .01$). Mean ratings for courses taken by Asian students were significantly lower than those of white, black, and Hispanic students (by 2.00, 1.91, and 1.58 percentage points, respectively).

In the regression analysis conducted to examine the impact of INTEREST ratings on graduation, the -2LL reduced by 1.98 from block one to block two; the omnibus test was significant ($p < .01$); and the Nagelkerke $R^2$ value was .109, suggesting that this predictor improved the model fit only slightly. The HL goodness-of-fit statistic was significant ($p = .012$), and after controlling for the effects of students’ race/ethnicity, SAT, income, and total GPA, the INTEREST variable was not a statistically significant predictor of graduation ($p = .158$).

Table 8

<table>
<thead>
<tr>
<th></th>
<th>$b$</th>
<th>SE($b$)</th>
<th>Wald $\chi^2$</th>
<th>Odds Ratio</th>
<th>-2LL</th>
<th>Model $\chi^2$</th>
<th>HL</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2996.90</td>
<td>223.02**</td>
<td>$p = .015$</td>
<td>.108</td>
</tr>
<tr>
<td>Interest</td>
<td>0.008</td>
<td>0.006</td>
<td>1.993</td>
<td>1.008</td>
<td>2994.92</td>
<td>225.00**</td>
<td>$p = .012$</td>
<td>.109</td>
</tr>
<tr>
<td>interest$\times$GPA</td>
<td>0.038</td>
<td>0.009</td>
<td>18.67**</td>
<td>1.038</td>
<td>2975.41</td>
<td>244.51**</td>
<td>$p &lt; .01$</td>
<td>.118</td>
</tr>
</tbody>
</table>

Note. *$p < .05$, **$p < .01$. Predictors are not shown. Refer to Table 6 for detailed baseline data.
As was done with GLOBAL the interaction between INTEREST and students’ race/ethnicity, as well as that between INTEREST and students’ total grade-point-averages, were examined by adding these variables to the logistic regression estimates in block three. The only interaction to reach statistical significance was that between students’ INTEREST course ratings and their total grade-point averages (Details are reported in Table 8). Adding the interest\textsuperscript{\mbox{}\times}\text{totalGPA} interaction term to the model resulted in a change in the -2LL from 2994.92 to 2975.41; the omnibus test was significant \( p < .01 \); and the Nagelkerke \( R^2 \) value was .118. Although the HL statistic was significant, these results suggest that the relationship between INTEREST ratings and graduation varies based on students’ total GPAs. The positive coefficient indicates that the relationship between this interaction and graduation is positive, and the odds ratio of 1.038 suggests that students’ odds of being retained increase by 3.8% for every one unit increase in the interaction between total GPA and INTEREST. As was the case with GLOBAL, students with high GPAs especially benefitted (in terms of their ability to graduate) from being enrolled in courses with high INTEREST ratings, that is to say, courses in which most students affirm that the instructor stimulates interest in the course.

It is also interesting to note that upon the addition of interest\textsuperscript{\mbox{}\times}\text{totalGPA} interaction term, the odds ratio of total GPA decreased from 2.811 to .153, suggesting that the addition of this interaction diminished the relationship between total GPA and persistence.

**Availability.** When determining how students’ course AVAILABILITY ratings related to the outcome variable as well as the other independent variables, the \( t \)-test conducted to gauge their relationship with graduation was not statistically significant. In other words, graduation did not vary based on students’ mean course assessments of
whether their instructors were available and willing to meet with them. AVAILABILITY ratings were significantly correlated with students’ high school GPAs ($r(3797) = -0.076$, $p < .01$) and SAT scores ($r(4217) = -0.086$, $p < .01$). AVAILABILITY also varied as a function of race/ethnicity ($F(4,4223) = 3.83$, $p < .01$). As with INTEREST, the AVAILABILITY ratings of courses taken by Asian students were lower: 1.11 percentage points lower than those of white students and 1.23 percentage points lower than those of Hispanic students.

The regression analysis for the AVAILABILITY variable indicated that the addition of this measure resulted in a decrease of 0.289 in the -2loglikelihood; the omnibus test was significant ($p < .01$); and the Nagelkerke $R^2$ value was .108—indicating that AVAILABILITY slightly improved the model’s ability to predict graduation. However, the HL statistic was still significant ($p = .017$), and after controlling for the effects of students’ race/ethnicity, high school GPA, income, SAT scores, FA, and total GPA, the AVAILABILITY variable was not a statistically significant predictor of graduation ($p = .591$).

Table 9

*Summary of Logistic Regression Model to Predict Impact of AVAILABILITY Student Ratings on Graduation*

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>SE(b)</th>
<th>Wald $\chi^2$</th>
<th>Odds Ratio</th>
<th>-2LL</th>
<th>Model $\chi^2$</th>
<th>HL</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td>2996.90</td>
<td>223.02**</td>
<td>$p = .015$</td>
<td>.108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability</td>
<td>-0.005</td>
<td>0.008</td>
<td>0.289</td>
<td>0.995</td>
<td>2996.61</td>
<td>223.31**</td>
<td>$p = .017$</td>
<td>.108</td>
</tr>
<tr>
<td>availGPA</td>
<td>0.041</td>
<td>0.013</td>
<td>10.07**</td>
<td>1.041</td>
<td>2986.34</td>
<td>233.58**</td>
<td>$p = .100$</td>
<td>.113</td>
</tr>
</tbody>
</table>

*Note. *$p < .05$, **$p < .01$. Predictors are not shown. Refer to Table 6 for detailed baseline data.*
The only interaction to reach statistical significance was that between students’ AVAILABILITY course ratings and their total grade-point averages (Details in Table 9). Including this interaction term in the model resulted in a reduction in -2LL of 10.275 (from 2996.614 to 2986.339); the omnibus test was significant (\( p < .01 \)); and the Nagelkerke’s \( R^2 \) value was .113. All of this suggests that the relationship between AVAILABILITY ratings and graduation varies based on students’ total GPAs. The positive coefficient indicates that the relationship between this interaction and graduation is positive, and the odds ratio of 1.041 suggests that students’ odds of being retained increase by 4.1% for every one unit increase in the interaction between total GPA and AVAILABILITY. Echoing the results for GLOBAL and INTEREST, students with high GPAs especially benefitted (in terms of their persistence) from being enrolled in courses with high AVAILABILITY ratings, namely courses in which instructors were available and willing to meet with students.

**Presentation.** Results for the final ratings variable—PRESENTATION—were surprising and somewhat counterintuitive. When first considering its relationship to students’ ability to graduate within 6 years, the \( t \)-test discerned differences in graduation associated with PRESENTATION \( t(1141)= 2.199, p < .05 \), with a Cohen’s \( d \) effect size of 0.09. In contrast to expectations and results for the other ratings variables, however, the mean PRESENTATION ratings of courses taken by retained students were slightly lower (by .62 percentage points) than those taken by students not retained. This suggests that students enrolled in classes in which the collective view was that the instructor presented course material effectively were less likely to be retained.
PRESENTATION was significantly correlated with students’ high school GPAs ($r(3797) = -0.098, p < .01$), SAT scores ($r(4217) = -0.125, p < .01$), and family income ($r(3797) = -0.039, p < .05$). They also varied based on students’ gender ($t(3665) = -3.127, p < .01$). Female students’ mean course ratings were only slightly higher than males’, however, as confirmed by the Cohen’s effect size value of -0.03. The addition of the PRESENTATION variable to the regression model containing the predetermined covariates resulted in a decrease in the -2LL of 7.45, suggesting that this predictor improved the model fit, a notion reinforced by the Nagelkerke $R^2$ value of .111.

Table 10

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>SE(b)</th>
<th>Wald $\chi^2$</th>
<th>Odds Ratio</th>
<th>-2LL</th>
<th>Model $\chi^2$</th>
<th>HL</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2996.90</td>
<td>223.02**</td>
<td>$p = .015$</td>
<td>.108</td>
</tr>
<tr>
<td>Present.</td>
<td>-0.021</td>
<td>0.008</td>
<td>7.31**</td>
<td>0.980</td>
<td>2989.45</td>
<td>230.47**</td>
<td>$p &lt; .01$</td>
<td>.111</td>
</tr>
<tr>
<td>present x GPA</td>
<td>0.032</td>
<td>0.012</td>
<td>7.25**</td>
<td>1.033</td>
<td>2982.20</td>
<td>237.72**</td>
<td>$p = .013$</td>
<td>.115</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01. Predictors are not shown. Refer to Table 6 for detailed baseline data.

Although the HL goodness-of-fit statistic was significant ($p < .01$), the PRESENTATION variable was a statistically significant predictor of graduation ($p < .01$) after controlling for the effects of students’ race/ethnicity, SAT, income, and total GPA. Since the coefficient is negative (-0.021), the odds ratio of .980 indicates that for every one
percentage point increase in course PRESENTATION ratings, a student’s odds of graduating diminish by 2.00%.

Finally, analyses were conducted to examine the effects of the interactions between PRESENTATION and students’ race/ethnicity as well as between PRESENTATION and students’ total GPA on graduation. Once again, the only interaction to reach statistical significance was that between students’ PRESENTATION course ratings and their total grade-point averages (Details are in Table 10). Including this interaction term in the model resulted in a reduction in -2LL of 7.25 (from 2989.45 to 2982.20), and the omnibus test was significant ($p < .01$). These results signal that this interaction term improves the model’s predictive capability, confirmed by the non-significant HL statistic ($p = .013$). All of this suggests that the relationship between PRESENTATION ratings and graduation varies based on students’ total GPAs. The positive coefficient indicates that the relationship between this interaction and graduation is positive—which contrasts results for the main effect of PRESENTATION. The odds ratio of 1.033 suggests that students’ odds of being retained increase by 3.3% for every one unit increase in the interaction between total GPA and PRESENTATION. As was the case for the other instructor behaviors/skills examined, students with high GPAs especially benefitted (in terms of their persistence) from being enrolled in courses with high PRESENTATION ratings, that is, courses in which most students felt that the instructor effectively presented the material.
Chapter Five: Summary, Conclusions, Discussion, and Implications

This chapter begins with a summary of the study, including its rationale and research questions, followed by conclusions that can be drawn from the results and a discussion of the findings. It then provides practical implications as well as recommendations for research based on both the study’s limitations and its results.

Summary

Nearly 40 years have passed since Tinto imparted his paradigmatic theory of student retention, thereby laying a foundation for much of the literature that followed and spurring countless researchers into action. Today, retention research efforts have gained momentum from increasing demands for accountability (Webber & Boehmer, 2008), rooted in the fact that degree completion is a problem at most U.S. institutions of higher learning, particularly for students from underrepresented racial and ethnic groups (Bowen, Chingos, & McPherson, 2009). At the same time, a new strand of retention studies has begun to investigate the classroom, an arena long protected from scrutiny. Retention literature has instead focused on such factors as students’ ability to become incorporated into an institution’s social and academic realms (Tinto, 1975, 1987, 1993), as well as the extent to which they become involved on campus (Astin, 1975, 1999). Most recently, Terenzini and Reason (2005) proposed that retention is the outcome of interconnected variables which fall into four major categories: student precollege characteristics and experiences, the organizational context, the student peer environment, and the individual student experience.

In keeping with these conceptual frameworks, empirical research has analyzed students’ precollege characteristics and experiences, as well as their out-of-class/co-
curricular experiences. Among the most potent factors that have been identified by scholars as associated with persistence are students’ race/ethnicity, socioeconomic levels, high school grade-point averages, and standardized test scores, as well as where they reside during their first year. The relatively new emphasis on the impact of faculty members and their pedagogies reflects a paradigm shift toward learner-centeredness and a growing awareness that for many of today’s students, the time spent in the classroom is the only regular contact with the institution. This being the case, the nature of students’ classroom experiences and their interactions with faculty members are thought to bear more on persistence than has been recognized (Reason, 2009; Tinto, 2006-7), and scholars have taken note that “there has been very little written about the importance of instruction and its impact on retention” (Giaquinto, 2009-10, p. 268).

Several recent studies suggest that faculty members’ pedagogies, particularly the use of active learning strategies, influence retention (Braxton, Milem, & Sullivan, 2000). Some evidence has also been marshaled suggesting that students’ perceptions of their instructors’ teaching skills are significant predictors of their chances of reenrolling and graduating (Braxton, Bray, & Berger, 2000; Pascarella, Salisbury, & Blaich, 2011; Pascarella, Seifert, & Whitt, 2008). The current study sought to extend this research vein. As opposed to collecting new data through surveys, as most previous studies have done, however, this investigation utilized student ratings of their professors. These data are already collected systematically on most college campuses and tend to reflect students’ overall views of their instructors as well as appraisals of instructors’ pedagogical behaviors, so they were presumed to be a potential (and possibly overlooked) source of insight into the departure puzzle.
Moving now from the theory to the practice of retention, it is worth reiterating that enrollment managers are increasingly held responsible for monitoring and improving retention rates, and that a hallmark of the practice of enrollment management is that it is a data-driven process (Hossler & Kalsbeek, 2008). Although enrollment managers have, for the most part, steered clear of institutions’ academic realms, top scholar-practitioners of enrollment management are now arguing that institutions cannot improve student retention without venturing into classrooms (Henderson, 2005; Kalsbeek, 2010). This study therefore explored the possibility that student ratings data could complement and enhance the practice of enrollment management and therefore assist institutions in achieving their student enrollment goals.

On the subject of student ratings themselves, the extensive body of literature on the subject (also known commonly as faculty and/or student evaluations) dispels myths that ratings are a “personality contest” and that high ratings can be secured by “giving” students high grades (Ory, 2001). Indeed, scholars affirm that the data they amass can provide crucial insights about instructor efficacy that have enormous potential for use in improving teaching practices and enhancing student learning (Theall & Franklin, 2000). Since ratings provide one measure of students’ perceptions of their instructors, this study envisioned an additional use for them—one based on the extent to which these data predict a student’s ability to graduate within six years. Because it was not possible to link an individual student’s ratings of his/her professors to his/her persistence, mean course ratings were utilized, such that the ratings variables came to represent one aspect of the academic context of students’ experiences. Regression analyses were used to explore the four specific research questions posed about the extent to which class average ratings
(both global and those for specific pedagogical behaviors) predict a student’s likelihood of graduating within six years, and the extent to which this effect varies for students from different racial/ethnic groups and/or with varying total grade point averages.

**Findings.** The first noteworthy finding was the positive relationship discerned between the mean course ratings of students in this sample and their retention. Specifically, students had enhanced odds of graduating if they enrolled in courses that received relatively higher GLOBAL ratings—those ratings that denote whether students’ overall evaluation of the instructor was positive and whether they would recommend this instructor to a friend. Students’ chances of persisting were not significantly influenced by either mean course ratings relating of instructors’ ability to generate interest in the course or their availability and willingness to meet with students. This study also bore the unexpected and seemingly counterintuitive finding that students’ odds of being retained were reduced as faculty members’ perceived presentation skills improved.

In response to the conjecture that the impact of ratings on retention varies for different groups of students, results indicated that this is, in fact, the case. The association between ratings and retention was significantly stronger for students with high total college GPAs. This was true for the overall instructor appraisal—i.e., GLOBAL—as well as for the three pedagogical variables examined: INTEREST, AVAILABILITY, and PRESENTATION. Interestingly, the impact of ratings on retention was comparable for students of different races and ethnicities. Regression analyses also found that college grade-point average was a crucial predictor of retention, and that household income was a significant predictor as well. In addition, the amount of financial aid offered to a student—in terms of the proportion of his/her financial need that was met—impacted
student persistence. To summarize, students’ odds of graduating improved commensurate with increases in their grade-point averages, with increases in their family incomes, and when greater proportions of their financial need were met. Findings associated with the retention of students from underrepresented minority groups ran counter to evidence that Hispanic and black students have lower odds of graduating within 6 years, simply by virtue of their minority status. In this sample, after controlling for the effects of students’ high school GPA, income, SAT scores, FA, and total GPA, no significant differences were identified in the graduation rates of students from different races or ethnicities.

Several results of the preliminary analyses conducted for the purpose of model building were also remarkable, especially those examining the relationship between student ratings and the other predictors. GLOBAL ratings were comparable for males and females, and these ratings were also similar for students of varied racial and ethnic backgrounds. GLOBAL ratings were significantly correlated with total GPA, yet both the GLOBAL and “total GPA” variables were each found to be significant predictors of graduation, even after controlling for the influence of the other. In the case of INTEREST and AVAILABILITY, variations were detected based on students’ races/ethnicities. Although this analysis did not control for the impact of other factors, it is worth noting that Asian students enrolled in more courses taught by instructors who received lower ratings on their ability to generate interest in the material as well as on their willingness and availability to meet with students.

Conclusions. Perhaps the most overarching conclusions that can be drawn from the study are that faculty members play a significant role in college student persistence and that student ratings data can provide useful insight about student retention. With
respect to ratings data, it appears that \textit{overall} mean course instructor ratings are more useful in predicting persistence than ratings of specific instructor pedagogical behaviors/skills—at least the ones included in this study. Additionally, class-level perceptions of an instructor’s ability to generate interest, of his/her willingness to meet with students, and of his/her presentation skills are important factors that improve retention for students with high grade-point averages. Overall, college grade-point average retains its position as a chief predictor of retention, and the utility of household income as a predictor of retention is upheld as well. One last conclusion that can be drawn from the results, one that contributes to a relatively under-examined category of retention research, concerns the role of financial aid in student persistence: The amount of financial aid offered to students—in terms of the proportion of financial need that is met—plays a significant role in their persistence.

\textbf{Discussion}

It would be remiss to begin discussing the results and conclusions of the current study without first acknowledging the countless scholars who have so diligently scrutinized the departure puzzle, as this study was conceptually rooted in and provides additional evidence for many of their theoretical and empirical tenets. For one, the positive associations between students’ mean course ratings and their persistence upholds the centrality of students’ academic integration to their subsequent persistence. If, as Ishitani and DesJardins (2002-3) have proposed, academic integration reflects students’ perceptions of their academic experiences and activities, as well as their perception of faculty concern for teaching (Braxton, Milem, & Sullivan, 2000), then the results of this study strengthen Tinto’s precept that academic integration is a major determinant of
student persistence. Along the same lines, the fact that two of the student pre-college characteristics and experiences examined (income and SAT), one marker of the organizational context (i.e., the first financial aid variable), one individual student experience (total GPA), and one classroom-related variable (mean GLOBAL ratings) were significant predictors of retention supports Terenzini and Reason’s (2005) comprehensive framework (See Figure 1) and their insistence that retention results from the interplay of various interconnected variables.

In turn, the fact that students’ overall course ratings were associated with enhanced odds of persisting resonates well with and extends the relatively new body of retention scholarship that explores the college classroom. This study provides additional evidence that faculty members are critical agents in the departure puzzle and that their teaching practices play a significant role in student success, in keeping with the findings of Braxton, Bray, and Berger (2000); Braxton, Milem, and Sullivan (2000); Lundquist, Spalding, and Landrum (2002); Pascarella, Salisbury, and Blaich (2011); and Pascarella, Seifert, and Whitt (2008). In contrast to Langbein and Snider’s (1999) finding that students’ ratings are positively associated with student persistence only up to a certain optimal level, after which a student’s odds of retention begin to decline, this study suggests that as students’ GLOBAL course ratings increase, so too do their odds of graduating. The unexpected finding that increases in mean course PRESENTATION ratings were associated with reduced odds of graduation is both puzzling and inconsistent with previous research: Presentation skills are often seen as chief indicators of effective college teaching (Lowman, 1995), and as described above, Braxton, Bray, and Berger (2000) found that students’ views of their instructor’s presentation skills were positively
associated with their persistence. The anomalous finding may reflect the myriad ways students can conceive of “presentation skills” as they complete their instructor ratings, and the absence of response process validity evidence makes it impossible to know how students interpret this construct.

Other results pertaining to student ratings are likewise worth discussing, as they inform the body of scholarship on faculty evaluations. The fact that only the GLOBAL ratings were significant predictors of retention when all of the other factors were considered is consistent with ratings scholars’ suggestion that these broad ratings items can be quite useful. Cashin and Downey (1992) and d’Apollonia and Abrami (1997) have asserted that questions about the overall effectiveness of the course and instructor tend to be good predictors of the overall quality of the course/instructor. It now appears they may also be good predictors of student retention. The rating behaviors of students from different racial/ethnic backgrounds in this study are also intriguing because these behaviors have not been expressly examined in published literature. Although the current study’s use of course-level ratings variables makes it difficult to draw conclusions about any racial or ethnic group, it is still peculiar that courses taken by white, Hispanic, and black students were rated in relatively similar ways, while those taken by Asian students received significantly lower ratings for instructors’ ability to generate interest and their availability. This could reflect disciplinary distinctions in the courses taken by Asian students, cultural causes, and/or the fact that Asian students are underrepresented at UM.

The results of interactions examined in this study are equally enigmatic. They beg the question: Why do students with high GPAs benefit from being enrolled in classes with higher ratings? One might instinctively presume that this finding reflects and/or
corroborates the myth that instructors “give” high grades in order to secure higher ratings. Upon closer examination, however, this logic flounders. Since the ratings data utilized were a class-level factor, they include the views of all students, not just those who had earned high grades. A more likely possibility is that this result further reinforces the potency of grades in student persistence (Recall that for every one unit increase in total GPA, a student’s odds of persisting were enhanced by 187%). In other words, the influence of mean course ratings on an individual student’s experience may be more pronounced when a student has already secured tangible evidence of his/her academic integration in the form of a high grade. This suggests that a high GPA may function as a precursor or prerequisite that must be met before the impact of other less powerful factors on retention can manifest itself. These results serve as important reminders, then, that instructors influence student retention both through their (seemingly innocuous) grading practices, and albeit to a lesser extent, through students’ perceptions of their teaching.19

One last set of findings worth deliberating pertains to the experiences of the Hispanic students in the sample. As has been mentioned, Hispanics are the fastest-growing minority group in the U.S., and the enrollment of Hispanic students at UM facilitated a comparison of their experiences with those of white students. Findings suggest that when students’ entering characteristics are similar, their likelihood of graduating is similar—regardless of their race/ethnicity; an encouraging discovery, indeed. Moreover, these results corroborate previous findings that students who first

19 Some could argue that a student’s grades and, by extension, grade-point average is an attribute of the individual student, yet as Hu (2005) has shown, college students’ grades are influenced by a variety of factors, including characteristics of students themselves, faculty, contextual measures of a course, discipline, and the institution. He reminds us that, in the absence of direct measures of student performance, “faculty members have the primary responsibility in…assigning grades to students” (p. 12).
enroll at minority-serving institutions are more likely to graduate (Pascarella & Terenzini, 2005). Although this may seem intuitive in the case of historically-black colleges and universities (HBCUs), which were established to serve and empower black students, the designation Hispanic-serving Institution (HSI) is a governmental classification based on an institution’s structural diversity. The fact that enrollment at an HSI may reap benefits similar to enrollment at an HBCU is thus notable, especially when one considers that about 100 colleges are poised to obtain the HSI designation (Santiago & Andrade, 2010).

By the same token, study findings cannot be understood without consideration of the larger context—the city of Miami. The percentage of Hispanic undergraduates at UM may itself positively influence Hispanic student achievement, yet it is also likely that Miami’s diverse population and overall climate for diversity promotes persistence for these students. Hispanics comprise 65% of Miami’s population and have a great deal of political and economic influence in the city (Lynch, 2000). It would thus be worthwhile to determine whether specific institutional practices at UM promote Hispanic student success, whether Hispanic student success in UM reflects the greater Miami milieu, or whether both occur synergistically. In either case, Hispanic students’ ability to graduate at the same rate as their non-Hispanic peers calls into question Tinto’s notion that students must disassociate themselves from their home communities to become integrated and persist. In fact, the Hispanic students examined may have been successful precisely because they were able to maintain their bicultural identities.

---

20 The classification HSI was created by the government and policy makers to describe colleges and universities in which 25% or more of students identify as Hispanic (Santiago & Andrade, 2010).
**Practical Implications**

The most straightforward practical implication of this study is that results could be shared with UM faculty as evidence that their teaching practices and interactions with students have considerable consequences. As Braxton (2008) has noted, instructors tend to think retention efforts are purely administrative matters that are not their concern. Confronted with proof that a one-unit increase in students’ mean GLOBAL ratings results in increased odds of graduating of 5.7%, faculty members may be more prone to embrace their role promoting student retention. Beyond this mere evidence sharing, the more-substantive implications described below are that institutions should help faculty understand their ratings and improve their teaching accordingly, and should consider integrating ratings data into their existing enrollment management analyses.

**Implications for ratings data interpretation.** For many instructors, looking at raw numbers and percentages, as ratings data are typically reported, offers little insight. Before they can adjust their teaching, they must decipher these data, and institutions can help. For instance, the University of Massachusetts, Amherst (UMASS-Amherst) recently developed a guide for its faculty on how to use ratings results (Stassen, 2012). After explaining the goals of the ratings program, the document breaks down the “score sheet” given to faculty members. Two unique features of these sheets are course statistics (elective vs. requirement, students’ class levels, expected grades, etc.) and a comparison of the instructor’s ratings with three groups.21 Finally, in a section aptly entitled “So, What do my results mean?” the guide points out the teaching construct reflected by each ratings survey item (i.e., questions 7 and 8 concern “helpful feedback to students”), and

---

21 These comparison classes are one from the department, one from the college, and one from the campus at large, selected based on shared course characteristics (such as size).
advises faculty to use the “teaching improvement packets” (TIPs) the University has developed, which provide pedagogical strategies aligned with the ratings items. Producing similar written materials (both a ratings guide and teaching packets) would be a feasible and fruitful step for institutions like UM where there is no centralized teaching assistance available to faculty members.²²

**Implications for college instruction.** The second major implication of this study is that it can be used to promote professional development opportunities. Given that most instructors have received minimal training in pedagogy, they cannot be expected to improve their teaching unless they are given access to research on contemporary students, how they learn, and the most effective instructional strategies with which to promote learning. Certainly, some of this information can be disseminated in written form (as in the TIPs packets described above), yet other formats such as workshops and one-on-one consultations may provide more thorough feedback. Given that the scholarship of teaching and learning in college (known as SoTL) is expansive, two specific strands of this literature that seem especially relevant are summarized below. These concern student motivation and a generational perspective on teaching.

As used herein, “motivation refers to the personal investment that an individual has in reaching a desired state or outcome” (Ambrose et al., 2010, p. 68). Ambrose and colleagues explain that “student motivation generates, directs, and sustains what they do to learn” (p. 69), which is one of their seven research-based principles for smart

---

²² Many U.S. colleges and universities have professional development centers focused on teaching, but the University of Miami does not.
teaching. Since motivation is, of course, a psychological construct, faculty members could argue that it is irrelevant to them, that students’ psychosocial development is the purview of student affairs professionals. Teaching scholars would respond that our classrooms are far from purely intellectual spaces; they are complex interpersonal and psychological arenas to which students bring their anxieties, ambitions, misconceptions, confidence, and so forth, such that an instructor’s interpersonal abilities and concern for students are central to effective teaching (Lowman, 1995; Sorcinelli, 2002).

Instructional approaches used to increase student motivation follow naturally from Ambrose et al.’s assertion that motivation is highest “if a goal is valued and expectancies for success are positive and the environment is perceived to be supportive” (p. 72). Faculty members can help students see the value in course materials, activities, and assessments by explicitly discussing their value, making connections to students’ goals for the class, and by providing authentic tasks (Ambrose et al., 2010). The phrase “expectancies for success” means that students cannot be motivated unless they think success is attainable. Students’ expectancies for success are more likely to be positive if instructors provide clear and detailed expectations (as in rubrics), provide frequent formative feedback, and set an appropriate level of challenge (Ambrose et al., 2010).

Finally, cultivating a supportive environment requires getting to know students, conveying respect for them, and promoting interaction amongst them (Sorcinelli, 2002).

23 The other six are that “students’ prior knowledge can help or hinder learning;” “how students organize knowledge influences how they learn and apply what they know;” “to develop mastery, students must acquire component skills, practice integrating them, and know when to apply what they have learned;” “goal-directed practice coupled with targeted feedback are critical to learning;” “students’ current level of development interacts with the social, emotional, and intellectual climate of the course to impact learning;” and “to become self-directed learners, students must learn to assess the demands of the task, evaluate their own knowledge and skills, plan their approach, monitor their progress, and adjust their strategies as needed” (Ambrose et al., 2010).
In addition to learning how to motivate students, faculty would benefit from acquiring a generational understanding of their students. Generational analyses are those that describe the characteristics of a college “generation,” a peer group defined by both its demographics and key life events which are, in turn, linked to shared values and behaviors. Instructors should know that their Millennial students (those born between 1982-2002) have been described as special, sheltered, confident, conventional, team-oriented, achieving, and pressured (Strauss & Howe, 2000). Others stress that these students are technologically-oriented (Black, 2010), narcissistic (Campbell & Twenge, 2010), and overwhelmed (Kadison & DiGeronimo, 2004); and yet another distinguishing feature is that Millennials are more racially and ethnically diverse than any other college generation (Black, 2010; Strauss & Howe, 2000).

For a thorough account of strategies for working with Millennial students, I refer readers to Wilson and Gerber’s (2008) excellent piece How generational theory can improve teaching. Their four main pedagogical suggestions are “enhanced clarity of both course structure and assignments” given Millennials’ sheltered upbringing and achievement-orientation; “student participation in course design” given their confidence and consumerist tendencies; “pre-planned measures to reduce stress” given how pressured they feel; and “rigorous attention to the ethics of learning” (p. 29) given Millennials’ narcissistic inclinations. Black (2010), who refers to contemporary students as “digital natives” and to most faculty as “digital immigrants,” suggests that instructors should supplement face-to-face instruction with online activities and materials, as well as create learning situations that parallel aspects of gaming. There are no clear-cut strategies for teaching students whose racial, ethnic, and experiential backgrounds differ markedly
from one’s own, other than to internalize McKeachie’s (2006) counsel that we must make every effort to avoid the deficit model, “the view that inadequate performance from an ethnic person automatically means the student is academically deficient, unmotivated, uninterested, or poorly prepared” (p. 152). Instead, faculty should get to know their students, learn about their cultures and countries of origin, and vary teaching approaches in response to varied learning styles.

**Implications for enrollment management.** As has been mentioned, enrollment managers are frequently charged with improving student retention, and this study conjectured that student ratings results could become an additional source of data for the enrollment management process. What, then, might University of Miami’s enrollment managers do in light of this study’s findings? They could simply add these data to their already expansive data sets where they collect and analyze student data, and to do so effectively would require the additional entry of data about each faculty member (his/her rank, for instance) as well as individual course data (whether elective or required, the course level, etc.), as well as the discipline since student ratings have been shown to vary based on these features (Ory, 2001).

It may be tempting to conduct surveys similar to the student ratings forms that can be connected to each individual student, so as to track his/her students’ progress. The results of this study suggest, however, that using existing ratings data is a cost effective alternative worth considering and tweaking for the purpose of identifying students at risk for departure. This seems especially true for UM because the student profile has already improved in the years subsequent to the cohorts examined. Research has shown that students enrolled in the most selective institutions earn the highest grades (Rojstaczer &
Healy, 2012). Taken together with the finding that student ratings were especially important to the retention of students with high GPAs, this research suggests that the predictive power of student ratings may have already increased and will continue to rise. The very same enrollment management team which has helped UM attract these high-achieving students can be instrumental in helping the institution retain them, and student ratings data show promise for use in these endeavors.

**Recommendations for Future Research**

Recommendations for future research based on the limitations of the current study will be presented first, followed by suggestions for future research based on the study’s findings.

**Limitations.** One considerable limitation of the current study was described previously—namely, that it examined the results of a group-level factor (mean course ratings) on an individual-level outcome (students’ graduation within 6 years) since the student ratings data used were collected anonymously. The study was able to reconcile this limitation, however, by using the class-level view of instructors as a contextual academic variable and interpreting results as class-level impacts. One way to bypass this obstacle altogether is to conduct separate surveys of students, i.e., not as part of a University-wide ratings system, and then track their progress throughout their college years, as Pascarella, Seifert, and Whitt (2008); and Pascarella, Salisbury, and Blaich (2011) have done. An alternative is to heed Kozub’s (2008) suggestion and adjust the entire student ratings system so that it is confidential instead of anonymous. This would allow institutions to use all of the data they can amass for individual students, including their actual grades as well as their actual ratings, to predict their odds of persistence and
devise appropriate interventions. Given that this proposition is contentious, however, an institution should first pilot this approach with a random sample of students and only move forward with the change if ratings are found to have a significant and sizable impact on retention.

Two of the study’s methodological limitations include the fact that only students’ affirmative ratings responses were utilized and that the amount of ratings data used per student was unequal. Future studies should attempt to secure or compute numerical values for ratings data that factor in students’ affirmative as well as negative responses, as well as any mid-level/neutral options, as this could increase the precision of findings. The decisions to use only UM’s core survey items and to include students in the study if at least 12 course ratings were available for them resulted in unequal numbers of ratings per student. To surmount this obstacle and facilitate University-wide comparisons across, the institution could mandate the use of a core set of items across all colleges and schools, and administering the ratings system online (as UM does currently) will facilitate the process as well since the data become more accessible.

One final set of limitations involves variables known to influence ratings and/or retention that were not included in the study. These include a disciplinary marker for students (as in their major, or at least STEM vs. non-STEM); a variable identifying where students resided, especially during their first year (i.e., in campus housing, at home with their parents, etc.); course characteristics (class size, whether required or elective, how many sections of same class were available); and instructor characteristics (race/ethnicity, rank). Also, some students are allowed to register for courses before their peers based on seniority and other factors (participation in athletics, for instance), and it appears that
students may self-select into courses with higher ratings when they are given access to University ratings data and can review online ratings (as in Ratemyprofessor.com). Including both of these factors—time of registration and the extent to which a student selected a class based on ratings—could thus be revealing. Given the complexity of the outcome at hand, incorporating more of these variables and factors in future studies would likely improve institutions’ ability to predict which of their students will and will not persist and secure a better understanding of retention on their campus.

In addition to these suggestions based on what the study did not do or could have executed more effectively, additional recommendations for future research spawn from the study’s actual findings. Now that UM has evidence that students’ overall, class-level instructor appraisals of their instructors are associated with their persistence, the University could investigate why this is the case and what occurs in its students’ minds as they complete these ratings. For instance, since class-level impressions of the pedagogical behaviors included in the core survey items examined did not prove useful in predicting student retention, the University could consider using alternative or supplemental ratings items. Indeed, a comprehensive psychometric evaluation of the current ratings system, especially as pertains to the validity of the ratings data it generates, would aid UM in making more informed decisions about how to improve faculty evaluation methods, in keeping with institution-specific goals and an institution-specific conception of “effective instruction.” Asking students to vocalize their thoughts while rating a course would also be a fruitful exercise which would provide validity evidence based on response processes, evidence that would begin to fill a void not only in the institution, but also in the canon of ratings literature more broadly.
Although not a limitation, per se, it is important to acknowledge that the results of this study are not generalizable since UM is distinctive in many important aspects. For one, its student body is not a representative sample of U.S. college students. These individuals were carefully selected and admitted to the institution based on some of the same variables examined herein (SAT scores and high-school GPA, for instance), variables known to be correlated with higher retention rates. The city of Miami is another distinctive feature whose influence on student retention should not be taken lightly, for reasons discussed above. On the other hand, the results of this study can perhaps prompt institutions with similar student profiles to conduct their own studies and determine the extent to which the ratings data they already collect predicts persistence for their students. UM, in turn, would appear to benefit from extending this research to future cohorts and from conducting complementary studies of first-to-second year persistence. It would also seem prudent to investigate the extent to which ratings for specific courses (required writing courses and/or gatekeeper introductory courses) are associated with students’ decision to reenroll for the second year.

In closing, it is important to credit the University of Miami for transforming itself, over the course of its only 87 years of existence, into a highly-selective and bustling hub of research and innovation that attracts students from all corners of the globe. The institution must also be commended for consistent improvements in first-to-second year persistence and graduation rates. In fact, 91% of the students who enrolled in 2010 re-enrolled for the second year, and 78% of students who enrolled in 2005 graduated within 6 years (Student Consumer Information, 2012). Even so, stakeholders feel the pressure and the responsibility to promote student persistence, as they recognize the increasingly
competitive nature of the postsecondary education market and the fact that UM’s graduation rates have played and will continue to play a key role in enhancing the institution’s visibility (Hurricane Sports, 2011). It is my hope that administrators and faculty alike will find reassurance in discovering that a UM student’s odds of degree completion are not predetermined, that the institution’s academic and administrative domains can collaborate to attain their shared goals of student learning and retention, and that student ratings data can contribute to these important endeavors.
References


Appendix A

Intended and Unintended Consequences of Using Student Ratings Data

**Intended**
- Instructors collect ratings, value the input, and make improvements in their teaching and courses.
- Instructors are rewarded for having excellent rating results (salary, promotion, awards, recognition).
- Instructors with very low ratings are encouraged by their department to seek help, possibly from colleagues or a campus faculty development office.
- Students perceive and use ratings as a means for indicating suggestions for improvement.
- Students have more information on which to base their course selections.
- Instructors use ratings as motivation to improve their teaching.
- Students perceive of ratings as a vehicle for change.

**Unintended**
- Instructors alter their teaching in order to receive high ratings (lower content difficulty, provide less content, give only high grades).
- The campus rewards poor teaching (lower faculty standards).
- Due to their convenience, the campus looks to student ratings as the only measure of teaching quality.
- The content of the student rating form may determine what is addressed in the classroom.
- Students reward poor teaching by believing they can give high ratings in return for high grades.
- Ratings are used to make discriminations between instructors that cannot be supported by the data.
- Due to the high stakes involved, instructors fail to follow proper administration procedures.
- The ratings process becomes a meaningless activity that is performed by students and instructors only because it is mandated.
Variable Definitions and Coding

Table B1

<table>
<thead>
<tr>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variables</strong></td>
</tr>
<tr>
<td>Graduation</td>
</tr>
<tr>
<td><strong>Independent variables</strong></td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>High-school GPA</td>
</tr>
<tr>
<td>Standardized test scores</td>
</tr>
<tr>
<td>Parents’ educational level (one per parent)</td>
</tr>
<tr>
<td>Financial aid a.</td>
</tr>
<tr>
<td>Financial aid b.</td>
</tr>
<tr>
<td>Income</td>
</tr>
<tr>
<td>College GPA 1st semester</td>
</tr>
<tr>
<td>College GPA total</td>
</tr>
<tr>
<td>Rating score</td>
</tr>
</tbody>
</table>
Appendix C

Table C1

*Number of Missing Cases per Variable*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Missing Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduation</td>
<td>0</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>132</td>
</tr>
<tr>
<td>High-school GPA</td>
<td>431</td>
</tr>
<tr>
<td>SAT</td>
<td>11</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>934</td>
</tr>
<tr>
<td>Father’s education</td>
<td>1135</td>
</tr>
<tr>
<td>Financial aid a.</td>
<td>804</td>
</tr>
<tr>
<td>Financial aid b.</td>
<td>804</td>
</tr>
<tr>
<td>Income</td>
<td>0</td>
</tr>
<tr>
<td>College GPA 1st semester</td>
<td>13</td>
</tr>
<tr>
<td>College GPA total</td>
<td>12</td>
</tr>
<tr>
<td>Global</td>
<td>11</td>
</tr>
<tr>
<td>Presentation</td>
<td>11</td>
</tr>
<tr>
<td>Interest</td>
<td>11</td>
</tr>
<tr>
<td>Availability</td>
<td>11</td>
</tr>
</tbody>
</table>
Appendix D

Results of Univariable Analyses

Table D1

Correlation Matrix among Continuous and Continuous Variables

<table>
<thead>
<tr>
<th></th>
<th>SAT</th>
<th>Income</th>
<th>FA_a</th>
<th>FA_b</th>
<th>1st GPA</th>
<th>Total GPA</th>
<th>Global</th>
<th>Presentation</th>
<th>Interest</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS GPA</td>
<td>.298**</td>
<td>-.089**</td>
<td>-.339**</td>
<td>.032</td>
<td>.314**</td>
<td>.307**</td>
<td>.002</td>
<td>-.098**</td>
<td>-.098**</td>
<td>-.076**</td>
</tr>
<tr>
<td>SAT</td>
<td>.129**</td>
<td>-.241**</td>
<td>.099**</td>
<td>.311**</td>
<td>.333**</td>
<td>.009</td>
<td>-.125**</td>
<td>-.124**</td>
<td>-.086**</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>-.158**</td>
<td>.142*</td>
<td>.061**</td>
<td>.093**</td>
<td>.000</td>
<td>-.039*</td>
<td>-.003</td>
<td>-.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA_a</td>
<td></td>
<td>-.140**</td>
<td>-.198**</td>
<td>-.182**</td>
<td>.051**</td>
<td>.035*</td>
<td>.076**</td>
<td>.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA_b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.056**</td>
<td>.060**</td>
<td>.008</td>
<td>.002</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>1st GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.792**</td>
<td>.150**</td>
<td>.005</td>
<td>.014</td>
<td>.002</td>
</tr>
<tr>
<td>Total GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.206**</td>
<td>.012</td>
<td>.086**</td>
<td>.034*</td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.572**</td>
<td>.616**</td>
<td>.501**</td>
</tr>
<tr>
<td>Presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.714**</td>
<td>.568**</td>
</tr>
<tr>
<td>Interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.561**</td>
</tr>
</tbody>
</table>

Note. **Correlation is significant at the 0.01 level.
Table D2

*Chi-square Values for Relationships among Categorical Independent Variables*

<table>
<thead>
<tr>
<th></th>
<th>Race/ethnicity</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>34.54**</td>
<td>4</td>
</tr>
<tr>
<td>Mother's education</td>
<td>105.17**</td>
<td>8</td>
</tr>
<tr>
<td>Father's education</td>
<td>85.34**</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note.* **Chi square is significant at the 0.01 level.*