6-25-2010

An Examination of the Predictive Validity of National Survey of Student Engagement Benchmarks and Scalelets

Amy M. Korzekwa

Amy M. Korzekwa
Candidate

Individual, Family and Community Education
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

, Chairperson

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF NATIONAL SURVEY OF STUDENT ENGAGEMENT BENCHMARKS AND SCALELETS

BY

AMY M. KORZEKWA

B.S., Psychology, University of New Mexico, 2007

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Arts
Educational Psychology

The University of New Mexico
Albuquerque, New Mexico

May, 2010

DEDICATION

This thesis is dedicated to my Theodore, for his love and support through this and everything else. It is because of you that I am here.

ACKNOWLEDGEMENTS

I would like to thank Dr. Scott Marley, my advisor and thesis chair, for his help and guidance throughout this project. I would also like to thank my committee members, Dr. Tom Root and Dr. Jay Parkes for their insights and guidance. Thank you to Vicky Dueer for her assistance with the data and encouragement in general. For help with the data analysis, I would like to thank Dr. James Selig. And lastly, thank you to my friends and family, for the unconditional support along the way.

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF NATIONAL SURVEY OF STUDENT ENGAGEMENT BENCHMARKS AND SCALELETS

BY

AMY M. KORZEKWA

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Arts
Educational Psychology

The University of New Mexico
Albuquerque, New Mexico

May, 2010

AN EXAMINATION OF THE PREDICTIVE VALIDITY OF NATIONAL SURVEY OF STUDENT ENGAGEMENT BENCHMARKS AND SCALELETS

by

Amy M. Korzekwa

B.S., Psychology, University of New Mexico, 2007

M.A., Educational Psychology, University of New Mexico, 2010

ABSTRACT

Universities use the National Survey of Student Engagement (NSSE) to evaluate university quality. Student engagement is theorized to relate to student outcomes. Recently, the validity of the NSSE has been questioned, with studies finding limited predictive validity. This study investigates the relationship between NSSE subscale scores (benchmarks and scalelets) and student outcomes, specifically first-year college GPA and third semester retention status. Linear and logistic regressions were used to evaluate the relationships, controlling for high school GPA, gender, and race/ethnicity. The subscales were minimally predictive of first-year GPA, but not at all predictive of third semester retention status. Implications and future directions for research are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction and Literature Review

Universities have long struggled to find a way to assess the quality of undergraduate education. Many methods exist, such as rankings like those found in *U.S. News and World Report*, which are based on the resources and reputations of the institution (Pike, 2004), or on selectivity, faculty credentials, or library holdings (Kuh, 2003). Other methods are based on student performance on standardized tests (Carini, Kuh & Klein, 2006). However, these methods fail to consider the activities and behaviors of the students while attending a university. The purpose of this study is to examine the validity of one method of assessing educational quality that does consider students' self-reported behaviors, the National Survey of Student Engagement (NSSE). This study discusses the construct of student engagement, early attempts at measuring student engagement, and associated reliability and validity issues. It then discusses the NSSE, two subscales configurations of the NSSE (benchmarks and scalelets), and prior research relating engagement to student outcomes and university quality. After discussing these issues, a description of the study is provided, followed by results, and a discussion of the results.

The Construct of Student Engagement

The use of student engagement as a method of assessment is a relatively new method of assessing undergraduate education. Although measuring engagement does not directly measure student learning, it is based on the idea that engaging in educationally relevant activities positively influences student learning (Kuh, 2003). The construct of student engagement has been examined for more than 70 years, although the wording has

changed over time (Astin, 1993; Pascarella and Terenzini, 2005; Pace, 1980). In the 1930s, Tyler used "time on task" as a measure of student engagement. In the 1960s and 1970s, Pace studied quality of effort. Tinto (1993) looked at social and academic integration. In 1991, constructs associated with engagement began to be called "student engagement" (Kuh, Schuh, Whitt, and Associates, 1991).

There is not a commonly agreed upon definition of engagement. However, Kuh's (2003) definition of engagement is often cited as: "the time and energy students devote to educationally sound activities inside and outside of the classroom" (p. 25). An alternative definition, also by Kuh (2001), is: "Student engagement represents both the time and energy students invest in educationally purposeful activities and the effort institutions devote to using effective educational practices" (Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008, p. 542). According to Kuh (2001), the effective educational practices described in these definitions are primarily based on Chickering and Gamson's (1987) *Seven Principles for Good Practice in Undergraduate Education*. In this book, Chickering and Gamson described the seven following categories of good educational practices: (1) student-faculty contact; (2) cooperation among students; (3) active learning; (4) prompt feedback to students; (5) time on task; (6) high expectations, and (7) respect for diverse students and diverse ways of knowing. These activities have been found to directly influence student learning as well as student educational experiences (Kuh, 2001). In addition, student engagement research is based on the premise that engaged students will invest more time and energy in their studies and subsequently learn more (Kuh, 2003).

Early Measures of Student Engagement

Pace (1980) is credited with developing the first survey to measure student

engagement, the College Student Experiences Questionnaire (CSEQ). The CSEQ is based on "quality of effort." He found that students gained more from college in personal, social, intellectual and educational areas when they committed more time and energy to educationally purposeful tasks. From the 1970s through the 1990s, the CSEQ was the primary scale to measure engagement (Kuh, 2009). However, little research has been done on how well the survey assesses university quality. Ethington and Horn (2007) analyzed the relationship between quality of effort (engagement) items and outcome items (gains in college) on the CSEQ. They found that quality of effort has a significant relationship with students' academic performances. However, a study by Kuh, Pace and Vesper (1997) found that correlations between student activity items and self-reported gain were not significant. In addition, the survey was not popular with universities. The main problem with the CSEQ was that the survey was long and difficult to administer, which resulted in lower response rates (Kuh, 2009). These mixed results indicated that the CSEQ may not have been producing scores that measure exactly what they were intended to measure.

Measurement Reliability and Validity

A primary concern with any instrument is that it produces reliable and valid scores. The *Standards for Educational and Psychological Testing* (The *Standards*, 1999) define validity as "… the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (p. 9). Furthermore, "The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (p. 9). Construct validity is when there is evidence to indicate that an individual's scores on an instrument are to be interpreted as the

individual's standing on the psychological construct that the instrument is intended to measure (The *Standards*, 1999). Rather than show multiple types of validity, one should create an argument for validity. However, every validity argument should be supported by multiple sources of evidence. In the case of student engagement scores, this means that when the scores are interpreted to be measures of the quality of a university, there should be a variety of evidence to support this claim.

The *Standards* (1999) describe five different sources of validity evidence. The first is evidence based on test content, which refers to how well the themes and presentation of the items, as well as the procedures for administering and scoring the test, match the construct they are intended to measure. The second source of evidence is based on response processes, which refers to whether the test-taker is going through the intended mental processes while completing the task. The third source of evidence is based on internal structure, which refers to how well the relationship of items and components of a test matches the construct they are intended to measure. The fourth source of evidence is based on consequences of testing, which refers to whether the consequences of a test are due to what the test is intended to measure or to an unintended outcome.

The final source of evidence for validity, and of importance to this study, is relations to external measures. This means that test scores should relate to other variables in the way the construct is theorized to relate to those variables. In the case of engagement, a basic premise is that student engagement is directly and positively related to student learning. Therefore, higher levels of engagement should be positively related to student educational outcomes and retention. If so, an argument for the validity of the

engagement constructs can be made. First, however, engagement needs to be measured by an instrument that can collect engagement data without the shortcomings of the CSEQ.

The National Survey of Student Engagement

In the late 1990s, engagement researchers felt that student engagement was a valuable measure that could be improved to serve as a method for assessing university quality (Kuh, Hayek, Carini, Ouimet, Gonyea, and Kennedy, 2001). To fill this need, the National Survey of Student Engagement (NSSE) was developed by the Indiana University Center for Postsecondary Research (IUCPR) in 1999 (Kuh et al., 2001). It was initially supported by The Pew Charitable Trusts, but became self-sufficient in 2003 as more universities paid to participate (Kuh et al., 2001). The primary purpose was to develop a method that uses evidence of student learning and effective educational practice, as opposed to rankings based on reputations and resources, to assess the quality of a college (Kuh et al., 2001). Additional purposes included: finding areas where a university can improve the college experience (Chen, Gonyea, Sarraf, BrckaLorenz, Korkmaz, Lambert, Shoup & Williams, 2009); discovering and documenting effective educational practices (Kuh et al., 2001); and advocating for public acceptance and use of empirically based measures of collegiate quality (Kuh et al., 2001). The survey was piloted in 1999, and the first administration was in 2000 to 276 colleges and universities (Kuh et al., 2001). The NSSE has since become a popular way to assess a university, expanding the number of participating schools each year. In 2008, 772 schools participated in the NSSE (Kuh, 2009).

The items on the NSSE fall into five subscales (Kuh et al., 2001). The first asks

students about their participation in educationally purposeful activities, such as the amount of time spent studying, or if they participated or will participate in a learning community. The second category asks students about their university's requirements, such as the amount of reading and writing. The third category asks students about the college environment, such as the amount of support offered, or the general relationship with students or faculty. This category also includes two direct questions about student satisfaction. The fourth category asks background questions, such as age, gender, living situation, educational status and major. The last category asks students to estimate their personal and educational growth since they began college, in areas such as general knowledge, written communication skills, and vocational preparation (see Appendix A).

According to the *NSSE Technical and Norms Report* (Kuh et al., 2001), many of items on the NSSE had already been used in college student research surveys, such as the CSEQ. However, several changes have been made since the first administration. Most significantly, the items on the test were evaluated in several different ways and changes were made based on the information gathered. For example, Ouimet, Bunnage, Carini, Kuh, and Kennedy (2004) used student focus groups, cognitive interviews, and expert advice to gather content- and process-related validity evidence for the original NSSE items after the first administration of the NSSE in 2000. They revised the survey according to the recommendations of the focus groups, which agreed with the expert review. They then interviewed students about the revised version to assess whether students were processing the questions and responses as intended.

Porter (2009) further evaluated NSSE items for evidence of validity based on test content. He argued that college students are unable to accurately recall their own

behaviors, especially when the behaviors are mundane, as are most of the behaviors studied in the NSSE. In addition, errors in recall are often biased to show the student in a better light. He also found that many of the questions, in an attempt to make the survey more attractive and easier to complete, are unnecessarily vague, resulting in different understandings of the questions between different students.

The NSSE, like all surveys, uses self-report data. This type of research asks participants to answer questions about their thoughts, attitudes, feelings and behaviors. For many measures, self-report is the only feasible source of data. However, as the participant is the only source of data, self-report is vulnerable to problems of accuracy of measurement. Research has found that the biggest problems with self-report are when respondents are unable to answer the question truthfully (Wentland & Smith, 1993) or unwilling to answer the question truthfully (Aaker, Kumar, & Day, 1998). Kuh (2004) describes the 5 conditions that have been found to make self-reports more likely to be valid: (1) when the information is known to the respondents; (2) the questions are phrased clearly and unambiguously; (3) the questions refer to recent activities; (4) the respondents think the questions merit a serious and thoughtful response; and (5) answering the questions does not threaten, embarrass, or violate the privacy of the respondent or encourage the respondent to respond in socially desirable ways (Bradburn & Sudman, 1988; Brandt, 1958; Converse & Presser, 1989; DeNisi & Shaw, 1977; Hansford & Hattie, 1982; Laing, Swayer, & Noble 1989; Lowman & Williams, 1987; Pace, 1985; Pike, 1995). Kuh (2004) explains that "NSSE was intentionally designed to satisfy all these conditions."

Another concern of survey research is response rate, or the percentage of the

sample of students who are asked to take the survey who actually participate. The NSSE tends to have average response rates of about 40% (Kuh, 2001). The main concern of low response rates is non-response bias, meaning that those in the sample who participate are different in some important way from those who do not participate (Chen, et al, 2009). However, studies by IUCPR (2008) have found only trivial differences in characteristics and behaviors of respondents and non-respondents. Proportional representation is also a concern about response rate, in that the portion of the sample who responds to the survey should match demographically the population from which they were drawn. However, during analysis, weights are used to account for over-response from groups such as women and full-time students (Chen, et al, 2009). In addition, the IUCPR has made efforts to improve response rate, such as making the survey look less like a test, or including experimental items of interest (Kuh, 2001). They also recommend linking participation to registration or graduation (Kuh, 2001).

Porter (2006) questions the validity of surveys, based on their self-selective nature. He argues that despite using a random sample to determine which students will be offered an opportunity to participate, those who choose to participate are fundamentally different from those who choose not to participate, based solely on the fact that they do or do not choose to participate.

NSSE Benchmarks

The NSSE is often broken into subscale scores that are called benchmarks (Kuh et al., 2001). Forty-two of the items on the NSSE are grouped into five benchmarks of good educational practices: *Level of Academic Challenge* (LAC), *Active and Collaborative Learning* (ACL), *Student-Faculty Interaction* (SFI), *Supportive Campus Environment*

(SCE), and *Enriching Educational Experiences* (EEE). These benchmarks were developed by the IUCPR to form cohesive, self-evident concepts (Kuh, 2001). The benchmarks were designed using a combination of empirical and conceptual analyses (Kuh et al., 2001). They first used Principal Components Analysis with an oblique rotation, then Exploratory Factor Analysis with an oblique rotation on the same sample to confirm the findings. Theory was then used to make sense of the groupings (Kuh et al., 2001). The benchmarks are intended to represent educational practices that are meaningful to faculty and administrators, as well as understandable to people outside a university. They are also designed to allow comparisons across universities, as well as measure changes across years within a university (Kuh, 2004).

Benchmark Reliability and Validity. The NSSE benchmarks have been evaluated for evidence of score reliability and validity. Test-retest reliability was conducted in 2002 and 2005 on the five NSSE benchmarks. It found reliability coefficients ranging from .74 to .78 and .69 to .74, respectively (IUCPR, 2009). In 2002, a principal components analysis was used to evaluate the groupings of the items into benchmarks, using the 2002 national data. A factor analysis was used again to confirm the previous analysis. The findings confirmed the original groupings (Kuh, 2004).

However, LaNasa, Cabrera, and Trangsrud (2009) used a confirmatory factor analysis to analyze the validity of the NSSE benchmarks at a public, doctoral, research university in the Midwest. They found that a five benchmark (factor) structure was not supported. Furthermore, a follow-up exploratory factor analysis yielded a 9 factor model instead. Porter (2009) also pointed out attempts by other researchers to replicate the five dimension structure have failed to confirm the theorized structure of the benchmarks.

Pike's Scalelets

Pike (2004) created 12 scalelets as an alternative way to group the NSSE items. He defines a scalelet as such: "A scalelet consists of a set of survey questions related to a specific aspect of the educational experiences of a group of students" (Pike, 2006). He found his scalelets to have greater explanatory power and to provide greater detail than the NSSE Benchmarks. His goal was to be able to access a particular construct, in a way that individual items and the survey as a whole cannot. The scalelets are designed to find a balance between the broad generalizations of the whole survey, and the very specific inferences of the individual items. Pike's scalelets are: *Course Challenge* (CC), *Writing Experience* (Wr), *Higher-Order Thinking Skills* (HOTS), *Active Learning* (AL), *Collaborative Learning* (CL), *Course Interaction* (CI), *Out-of-Class Interaction* (OCI), *Varied Experiences* (VE), *Information Technology* (IT), *Diversity* (Div), and *Support of Student Success* (SSS). The items in each scalelet were chosen based on face and content validity (Pike, 2006). The content of the scalelets parallels the content of the NSSE benchmarks. However, the scalelets use 46 items and divide the content into a greater number of subscales. For example, the items of the Level of Academic Challenge benchmark are distributed among the Course Challenge, Writing Experiences, and Higher-Order Thinking Skills scalelets.

Scalelet Reliability and Validity. The reliability of the scalelets was evaluated by Pike (2006). He found reliability coefficients ranging from $E\rho^2 = 0.72$ to $E\rho^2 = 0.94$. Pike (2006) tested the validity of his scalelets and the NSSE benchmarks by looking for evidence in relationships with other variables. Specifically, he looked for evidence that the scalelets and benchmarks relate as expected to student outcomes, which he called

convergent validity, and for evidence that the scalelets and benchmarks relate differently to different outcomes, which he called discriminant validity. Using data from a 2004 administration of the NSSE, and using institution as the unit of analysis, he used four multiple regression models to test the relationships between NSSE benchmarks and general education gains and practical-skill gains (both gains measures are derived from items on the NSSE), and between NSSE scalelets and general education gains and practical-skill gains.

The analyses found that institutional characteristics and Pike's scalelets together account for 81.3% of the variance in general education gains with the scalelets accounting for 34.0% of unique variance in general education gains. This is in contrast to a regression with the same data using NSSE benchmarks, which found that institutional characteristics and the benchmarks together account for 78.0% of the variance in general education gains, while the benchmarks accounting for 30.7% of unique variance in general education gains. A similar set of analyses was conducted using practical-skill gains instead of general education gains, with more striking results. These analyses found that institutional characteristics and Pike's scalelets together account for 53.6% of the variance in practical-skill gains with Pike's scalelets accounting for 35.5% of unique variance in practical-skill gains. Institutional characteristics and the NSSE benchmarks together account for 40.3% of the variance in practical-skill gains, while the benchmarks account for 22.2% of unique variance in practical-skill gains.

Pike argues that the shared variance indicates that the scalelets and benchmarks are related to the outcomes, which is criterion-related validity evidence. In addition, the scalelets had a more differentiated relationship with the gains than the benchmarks,

showing discriminant validity. The benchmarks and scalelets also related differently to the general education gains than to the practical-skill gains, which was also evidence of discriminant validity.

NSSE and Educational Outcomes

The NSSE items, benchmarks, and scalelets have been found to relate to educational outcomes in many ways. A study (Carini, et al, 2006) of 2004 data from 1058 students at 14 universities found that individual items on the NSSE correlate positively with critical thinking and grades. A research study conducted by IUCPR (Kuh, Kinzie, Cruce, Shoup, & Gonyea, 2007) looked at data from about 6,000 first-year students and 5,000 senior students at 18 universities over four years (2000-2003) to investigate how student engagement, as measured by 19 items on the NSSE, relates to educational outcomes, such as first-year grade point average (GPA), persistence to the second year and senior GPA. They found that engagement has small, positive effects on both grades and persistence, for students of all racial and ethnic backgrounds, even after controlling for pre-college variables. In addition, these effects were found to be stronger in lower ability students and students of color. The NSSE benchmarks have been found to relate positively to learning outcomes as well. They have been found to relate strongly to both high levels of learning and personal development (Kuh, 2001). Kuh (2004) found that all five benchmarks correlate positively to student-reported GPA.

On the other hand, not all studies have found strong relationships between NSSE measures and student outcomes. Gordon, Ludlum and Hoey (2008) looked at the relationship between three forms of NSSE (benchmarks, scalelets, and individual items) and four student outcomes (freshman GPA, freshman retention, senior GPA, and senior

job placement) with a sample of students from the Georgia Institute of Technology. They found only a small relationship between the NSSE benchmarks and student outcomes, and a slightly better relationship between Pike's scalelets and student outcomes. Some of the individual items were found to be more related to the student outcomes than the benchmarks or scalelets, but still had mixed results.

Research has been conducted on how engagement relates to different student characteristics. A study (Kuh, 2003) of all of the data collected in the first three administrations of NSSE found some differences between groups in levels of engagement. They found that women, full-time students, students living on campus, native students, students who participate in learning communities, international students, and students with diversity experiences are all more likely to be engaged than their counterparts. Carini, et al. (2006) found that low-ability students are more likely to benefit from engagement. In addition, seniors and freshman benefit from different forms of engagement, and institutions vary in how well they convert engagement into higher performance.

All of this research on the relationship between engagement and educational outcomes helps to shed light on the research that has been done on group differences in educational outcomes. Kuh, et al. (2007) found that women, White students, and students with one or more college-educated parents have higher first-year GPAs in general over their counterparts. All of this becomes especially relevant when considering the Carini, et al. (2006) finding that engagement is most likely to be beneficial to low-ability students.

Thus far, the research on the relationship between the NSSE and student outcomes has been inconsistent. Although the NSSE creators have conducted many

studies to confirm this relationship, few researchers uninvolved in the project have been able to replicate their findings. In addition, many of the studies used student outcomes that were measured by the NSSE, rather than external variables. This situation requires further research to establish the relationship between NSSE and externally measured student outcomes.

Study Purpose

The purpose of the present study is to evaluate the predictive validity of NSSE subscale scores by investigating their relationship with external variables. This study investigated the relationship between the NSSE subscales and student outcomes at the University of New Mexico. More specifically, this study compared the relationships of the NSSE benchmarks and scalelets in relationship to student retention and first-year GPA. Additionally, the relationship between background variables (i.e., gender, race/ethnicity, and high school GPA) and student achievement and retention was examined. Based on these purposes the following research questions were addressed:

RQ1: How do the NSSE benchmarks relate to retention and first-year GPA?

RQ2: How do Pike's scalelets relate to retention and first-year GPA?

RQ3: Which technique of aggregating items into subscales results in greater criterion-related validity?

RQ4: How does gender relate to retention and first-year GPA?

RQ5: How does race/ethnicity relate to retention and first-year GPA?

RQ6: How does high school GPA relate to retention and first-year GPA?

RQ7: Does gender interact with the NSSE subscales in the prediction of retention and first-year GPA?

Chapter 2

Methodology

Sample

The population for this study is freshmen at the University of New Mexico (UNM). In Fall 2008, there were 4,154 students in the freshman class. Freshmen were defined as all students with fewer than 30 credit hours. In Spring 2009, 2,671 of those freshmen were randomly selected to participate in the study. Of this number, 619 freshmen responded for a response rate of 23.2%. All participating subjects provided informed consent as required by the UNM Institutional Review Board on Human Subjects.

Instrumentation

National Survey of Student Engagement. The NSSE consists of 85 questions in 13 categories (see Appendix A). Each question is answered on a Likert-type scale, with response scales ranging from 4 to 8 options. The items are divided into categories based on what sort of question is being asked, such as behavior or opinion, as well as what responses are used to answer the questions.

Benchmarks. The NSSE is often broken up into 5 benchmarks, used to capture vital aspects of the student experience. The benchmarks can be found in Appendix 2. The NSSE benchmarks were calculated using the Statistical Package for Social Sciences (SPSS) syntax found on the NSSE website (http://nsse.iub.edu/html/syntax_library.cfm). The calculations were done by transforming each item to a scale of 0 to 100. For example, an item that previously had response options 1 through 4 would have response options of 0, 33.3, 66.6 and 100. If the participant answered about 60% of the items in the

benchmark, then the item scores were averaged to create a benchmark score. Appendix 2 lists the items in each benchmark.

Pike's Scalelets. Pike (2004) developed 11 scalelets as another way of grouping the NSSE items. The SPSS syntax for creating these scalelets is available at the NSSE website (http://nsse.iub.edu/html/syntax_library.cfm). The scalelets were created by converting all responses to a scale of 0 to 100, in the same manner as the benchmarks. Again, if the participant has answered at least 3 out of 3, 3 out of 4, 4 out of 5, or 6 out of 9 of the items, the responses are averaged to form the scalelets score. Appendix 2 lists the items in each scalelet.

High School and College Grade Point Averages. College GPA was collected from the registrar, and included all complete semesters up to the analysis. It is measured on a scale of 0 to 4.33, with 0 representing a failing grade (F) in a class, and 4.33 representing the highest possible grade (A+). High school GPA was also collected from the registrar, and included the cumulative GPA from all four years of high school. It is measured on a scale of 0 to 4, with 0 representing a failing grade (F) and 4 representing an A.

Retention. Retention data was gathered from the registrar as well. Students who returned to UNM for their third semester in fall 2009 were considered "retained" while students who did not return were considered "not retained." Although many students may have transferred to a different school, rather than dropped out of school, we were interested in both success at and satisfaction with UNM.

Demographics. The NSSE includes a section of questions on demographics. The survey asks students for their age, gender, race/ethnicity, and estimate of grades in college (by letter grade: A, A-, B+, B, B-, C+, C, C- or lower). The gender and

race/ethnicity variables were also retrieved from the registrar to be confirmed. In the case of discrepancies, the registrar data was used. These demographics were used to account for external factors that might have an influence on college GPA, retention, or engagement.

Procedure

In February, 2009, the IUCPR sent all students in the sample emails with invitations to participate in the survey. The emails included information about the survey, a description of the incentives, and a link to the survey administration website. The first page of the website was a description of participation, followed by a description of the confidentiality procedures and who to contact with questions or concerns. Participants were then asked to choose whether to proceed with the survey or decline participation by clicking the appropriate button. The website was designed to allow students to begin the survey, then exit the website and return at a later time to complete it. Students in the sample were sent 4 additional emails to remind them to participate. The last reminder was sent in April, and the survey administration website closed June 1.

The IUCPR compiled the data and returned reports of frequencies and means, as well as the raw data sets. All responses were included in the data. The NSSE benchmarks and Pike's scalelets were calculated using the SPSS syntax on the NSSE website.

Analyses

The Statistical Package for the Social Sciences version 17.0 (SPSS) was used to analyze the data. First, variables were analyzed for differences between respondents and non-respondents. Second, Cronbach's alpha (Cronbach, 1951) was calculated for the benchmarks and scalelets. Third, raw and disattenuated correlations were calculated for

the benchmarks, scalelets, and GPA variables. Last, regression analyses were conducted. Multiple regression was used to predict college GPA (Pedhazur, 1997), while logistic regression was used to predict retention (Pampel, 2000). The analyses were performed for each outcome once using NSSE benchmarks and once using Pike's scalelets. Demographic information and high school GPA were statistically accounted for in all analyses.

Prior to the primary analyses, a statistical comparison of respondents and non-respondents was conducted to assess sampling bias. Students in the sample who participated in the survey were compared to students in the sample who did not participate in the survey on high school GPA, first-year college GPA, retention rates, race/ethnicity distribution, and gender distribution. T-tests were used for the GPA variables, and Chi-square tests were used for the race/ethnicity, gender and retention.

College GPA. The first analysis looked at the relationship between NSSE subscales and college GPA. First, the following statistical assumptions associated with multiple regression were assessed: 1) linearity of relationships, 2) normality of variables, 3) the variance of errors is the same across all levels of the independent variables (i.e., homoscedasticity: see Pedhazur, 1997). Linearity of relationships was assessed by viewing scatterplots between each independent variable and dependent variable. Normality was assessed by viewing histograms of each variable. Homoscedasticity was assessed by examining the scatterplot of the standardized residuals and the standardized predicted values. Correlation coefficients were calculated to determine the zero-order relationships between all of the variables. Disattenuated correlations were also calculated to determine the relationships between the variables, given perfect reliability.

Hierarchical multiple regressions were then conducted (Cohen, Cohen, Aiken and West, 2003). A block method of entry was used. The first regression was conducted to examine the relationship between demographic variables, NSSE benchmarks and the interactions of gender with NSSE benchmarks with college GPA. The first block of the model contained the demographic variables; the second block the NSSE benchmarks, and the third block interactions of the demographic characteristics with NSSE benchmarks. This allowed all of the demographic variables to account for as much variance in GPA as possible before the NSSE benchmarks were entered into the model. All benchmarks were mean-centered to aid in the interpretation of the interaction variables. The second regression was conducted to examine the relationship between demographic variables, Pike's scalelets, and college GPA. The steps were the same as with NSSE benchmarks, first entering the background characteristics, then the scalelets to see how much additional variance they accounted for. For both analyses a type I error rate of .05 was used to establish statistical significance. Significant predictors were interpreted using raw and standardized beta coefficients.

Retention. The second analysis looked at the relationship between NSSE and retention to the second year of college. First, the following statistical assumptions associated with logistic regression were assessed: 1) the dependent variable is dichotomous; 2) the independent variables are not multicollinear; 3) independence of error, and 4) proper model specification (Pedhazur, 1997). Multicollinearity of independent variables was assessed by looking at the correlations of the independent variables.

After assessing the assumptions, a hierarchical logistic regression was conducted.

The first block consisted of entering the demographic characteristics. The second block entered the NSSE benchmarks to see how much variance the benchmarks account for using pseudo R^2, after the background characteristics accounted for their share of the variance. A third block was entered that included the interactions of the demographic characteristics with NSSE benchmarks. The third block was assessed in the same manner as the second block. The second logistic regression was conducted in the same manner to examine the relationship between demographic variables, Pike's scalelets, and retention. For both analyses a type I error rate of .05 was used to establish statistical significance. Beta coefficients and odds ratios for statistically significant independent variables were interpreted.

Chapter 3

Results

Sample Characteristics

Respondents and non-respondents were compared on demographic characteristics, high school GPA, first-year college GPA, and retention rates (See Table 1). Students who responded to the survey (M = 3.45, SD = 0.47) had statistically higher high school GPAs than non-respondents (M = 3.27, SD = 0.48), $t(2019) = 7.41$, $p < .001$, Cohen's $d = .38$. Respondents also had higher first-year GPAs (M = 3.03, SD = 0.73) than non-respondents (M = 2.80, SD = 0.80), $t(2360) = 6.23$, $p < .001$, Cohen's $d = .29$. There was no difference in age between respondents (M = 19.09, SD = 2.90) and non-respondents (M = 19.05, SD = 2.20), $t(2635) = .329$, $p = .742$. A chi-square test of independence comparing the percentage of respondents who were female (67.0%) to the percentage of non-respondents who were female (51.5%) found a statistically significant difference X^2 (1, N = 2637) = 45.790, $p < .001$. A chi-square test of independence comparing the percentage of respondents who enrolled for a third semester (88.4%) to the percentage of non-respondents who enrolled for a third semester (71.3%) found a statistically significant difference X^2 (1, N = 2671) = 74.075, $p < .001$. A chi-square test of independence of race/ethnicity found that there were no statistically significant differences between students who responded and students who did not, X^2 (6, N = 2637) = 2.721, $p = .843$.

Benchmark and Scalelet Reliabilities

Cronbach's alpha (α) (Cronbach, 1951) was used to assess reliability of the benchmark and scalelet scores, and is reported in Tables 2 and 3. The benchmark reliabilities ranged from α = .63 to α = .74, with four out of five benchmarks at .70 or

Table 1: Respondent and Non-respondent comparisons

Variable	Respondents	Non-respondents
GPA:	**Mean**	
*High School**	3.45	3.27
*First-year cumulative**	3.03	2.80
Age:	19.05	19.09
Gender:	**Percent**	
*Females**	67.0%	51.5%
*Males**	33.0%	48.5%
Retention:	**Percent**	
*Enrolled**	88.4%	71.3%
*Not enrolled**	11.6%	28.7%
Ethnicity:	**Percent**	
White, non-Hispanic	43.2%	43.5%
Hispanic	38.2%	39.6%
American Indian	6.2%	4.8%
African American/Black	4.0%	3.6%
Asian/Pacific Islander	3.8%	3.9%
International	1.0%	1.5%
No Response	3.6%	3.3%

** Differences are statistically significant at $p < .001$.*

higher. The scalelet reliabilities ranged from α = .47 to α = .76, with three out of twelve scalelets at .70 or higher (See Tables 4 and 5).

Zero-Order Correlations (Observed and Disattenuated)

Zero-order correlations were calculated between benchmarks and the two GPA variables, and between scalelets and the two GPA variables (Tables 2 and 3). All of the benchmarks were significantly correlated with each other (*Pearson's r* = .308 to *r* = .606, $p < .001$), as were the scalelets (*r* = .150 to *r* = .584, $p < .001$). In addition, three of the benchmarks were significantly correlated with high school GPA (LAC: *r* = .104, *p* = .018, SFI: *r* = -.085, *p* = .042, EEE: *r* = .082, *p* = .048), and three were significantly correlated with first-year college GPA (LAC: *r* = .125, *p* = .006, ACL: *r* = .105, *p* = .016, EEE: *r* = .147, *p* = .001). Four of the scalelets were significantly correlated with high school GPA (CC: *r* = .096, *p* = .028, AL: *r* = .091, $p < .05$, CL: *r* = .086, *p* = .045, VE: *r* = .141, *p* = .003), while none were significantly correlated with first-year college GPA. High school GPA and first-year college GPA were statistically significantly correlated (*r* = .534, $p < .001$). Disattenuated correlation coefficients were also calculated (Tables 4 and 5).

Prediction of GPA using Benchmarks

The first multiple regression was conducted to look at the relationship between NSSE benchmarks and first-year college GPA. When assumptions were assessed, all variables were found to relate linearly to first-year GPA, to be reasonably normal, and to have similar levels of error across all levels of the variable. Only students with complete data were included in the analysis. In this case, there were 413 students in the analysis.

Table 2: Benchmark zero-order correlations and reliabilities

Variable	M	SD	*1*	*2*	*3*	*4*	*5*	*6*
1. Academic Challenge	52.76	12.82	(.698)					
2. Active and Collaborative Learning	43.56	16.12	.576**	(.696)				
3. Student-Faculty Interaction	34.23	18.33	.472**	.606**	(.719)			
4. Enriching Educational Experiences	29.53	13.09	.496**	.533**	.445**	(.631)		
5. Supportive Campus Environment	55.50	17.86	.400**	.358**	.409**	.348**	(.737)	
6. High School GPA	3.14	.66	.104*	.049	-.085*	.082*	-.030	
7. Cumulative First-Year GPA	3.47	.46	.125**	.105*	-.053	.147**	.008	.534*

Note: Reliabilities are on the diagonal. * p < .01, ** p < .05

Table 3: Scalelet zero-order correlations and reliabilities

Variable	M	SD	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Course Challenge	59.35	15.14	(.583)												
2. Writing	44.74	14.16	.340**	(.554)											
3. Active-Learning Experiences	41.20	19.46	.230**	.401**	(.559)										
4. Collaborative-Learning Experiences	45.08	18.19	.274**	.349**	.473**	(.539)									
5. Course-Related Interactions with Faculty	47.69	21.27	.327**	.395**	.506**	.480**	(.592)								
6. Out-of-Class Interactions with Faculty	20.87	20.02	.279**	.244**	.472**	.463**	.584**	(.525)							
7. Use of Information Technology	68.71	20.48	.225**	.278**	.277**	.244**	.329**	.189**	(.472)						
8. Emphasis on Diversity	56.96	24.97	.207**	.242**	.318**	.428**	.342**	.239**	.277**	(.667)					
9. Varied Educational Experiences	20.85	13.97	.217**	.215**	.434**	.281**	.233**	.392**	.204**	.224**	(.584)				
10. Support for Student Success	49.55	24.07	.288**	.172**	.277**	.273**	.303**	.358**	.219**	.302**	.287**	(.763)			
11. Interpersonal Environment	61.61	18.64	.363**	.150**	.173**	.273**	.234**	.279**	.173**	.198**	.239**	.378**	(.695)		
12. Higher Order Thinking Skills	66.87	18.71	.351**	.364**	.378**	.432**	.350**	.319**	.368**	.425**	.259**	.309**	.280**	(.749)	
13. Cumulative First-Year GPA	3.46	.47	.096*	.007	.091*	.086*	-.031	-.061	.033	.077	.141**	-.055	.065	.060	
14. High School GPA	3.14	.65	.049	-.037	.016	.056	-.076	-.061	.005	.056	.060	-.016	-.028	.056	.532**

Note: Reliabilities are on the diagonal. * $p < .05$, ** $p < .01$

Table 4: Benchmark disattenuated correlations

Variable	M	SD	*1*	*2*	*3*	*4*	*5*	*6*
1. Academic Challenge	52.76	12.82	1					
2. Active and Collaborative Learning	43.56	16.12	.826**	1				
3. Student-Faculty Interaction	34.23	18.33	.666**	.857**	1			
4. Enriching Educational Experiences	29.53	13.09	.747**	.804**	.661**	1		
5. Supportive Campus Environment	55.50	17.86	.558**	.500**	.562**	.510**	1	
6. High School GPA	3.14	.66	.124*	.059	-.100*	.103*	-.035	n/a
7. Cumulative First-Year GPA	3.47	.46	.150**	.126*	-.063	.185**	.009	n/a

* p < .05, ** p < .01

Table 5: Scalelet disattenuated Correlations

Variable	M	SD	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Course Challenge	59.35	15.14	1												
2. Writing	44.74	14.16	.598**	1											
3. Active-Learning	41.20	19.46	.403**	.721**	1										
4. Collaborative-Learning	45.08	18.19	.489**	.639**	.862**	1									
5. Course-Related Interactions	47.69	21.27	.557**	.690**	.880**	.850**	1								
6. Out-of-Class Interactions	20.87	20.02	.504**	.452**	.871**	.870**	1.048**	1							
7. Use of Information Technology	68.71	20.48	.429**	.544**	.539**	.484**	.622**	.380**	1						
8. Emphasis on Diversity	56.96	24.97	.332**	.398**	.521**	.714**	.544**	.404**	.494**	1					
9. Varied Educational Experiences	20.85	13.97	.372**	.378**	.760**	.501**	.396**	.708**	.389**	.359**	1				
10. Support for Student Success	49.55	24.07	.432**	.265**	.424**	.426**	.451**	.566**	.365**	.423**	.430**	1			
11. Interpersonal Environment	61.61	18.64	.570**	.242**	.278**	.446**	.365**	.462**	.302**	.291**	.375**	.519**	1		
12. Higher Order Thinking Skills	66.87	18.71	.531**	.565**	.584**	.680**	.526**	.509**	.619**	.601**	.392**	.409**	.388**	1	
13. Cumulative First-Year GPA	3.46	.47	.126*	.010	.122*	.117*	-.040	-.084	.047	.094	.185	-.063**	.078	.069	n/a
14. High School GPA	3.14	.65	.064	-.049	.021	.076	-.099	-.085	.007	.069	.079	-.018	-.033	.065	n/a

* p < .05, ** p < .01

The regression was constructed using a block method of entry. The first block of variables entered into the model contained high school GPA, gender (0 = male, 1 = female), and race/ethnicity (dummy-coded with White/non-Hispanic as the reference). The second block contained the five benchmarks. The third block contained the interactions between gender and the benchmarks (See Appendix C, Table 6).

The first block of variables accounted for 30.3% of the variance in first-year GPA, $F(8, 404) = 21.986$, $p < .001$, $R^2 = .303$. The overall regression for the second block of variables (the NSSE benchmarks) was statistically significant $F(13, 399) = 14.664$, $p < .001$, $R^2 = .323$. Furthermore, the change in variance accounted for was statistically significant, $\Delta F(5, 399) = 2.357$, $p = .040$, $\Delta R^2 = .020$. The third model, containing all three blocks of variables, accounted for 32.8% of the variance in first-year GPA, $F(18, 394) = 10.704$, $p < .001$, $R^2 = .328$. However, the increase in variance accounted for was not statistically significant, $\Delta F(5, 394) = .600$, $p = .700$, $\Delta R^2 = .005$. Since the third block containing the interactions does not significantly increase R^2, the second model containing demographic variables and the benchmarks was interpreted (see Cohen, Cohen, Aiken & West, 2003 for detailed rationale).

The second model had three significant coefficients. High school GPA was a significant positive predictor ($B = .702$, $p < .001$), and gender was a significant positive predictor ($B = .162$, $p = .007$), which indicated that females had higher GPA. The only benchmark that was statistically significant was Student-Faculty Interaction, which was a negative predictor of first-year GPA ($B = -.005$, $p = .022$).

Prediction of GPA using Scalelets

The regression was conducted in the same manner as the benchmarks regression,

using the scalelets in place of the benchmarks. Again, the first block contained high school GPA, gender (0 = male, 1 = female), and race/ethnicity (dummy-coded with White/non-Hispanic as the reference). The second block contained the twelve scalelets. The third block contained the interactions between gender and the scalelets (See Appendix C, Table 7). The analysis included the 387 students who had complete data.

The first block of variables accounted for 30.6% of the variance in first-year GPA, $F(8, 378) = 20.852$, $p < .001$, $R^2 = .306$. The overall regression for the second block of variables (the NSSE scalelets) was statistically significant $F(20, 366) = 9.761$, $p < .001$, $R^2 = .348$. This change was statistically significant, $\Delta F(12, 366) = 1.948$, $p = .028$, $\Delta R^2 = .042$. The third model, containing all three blocks of variables, accounted for 41.1% of the variance in first-year GPA, $F(32, 354) = 7.026$, $p < .001$, $R^2 = .411$. This increase in variance accounted for was statistically significant, $\Delta F(12, 354) = 1.958$, $p = .027$, $\Delta R^2 = .041$. Since the third block containing the interactions significantly increases R^2, the third model containing demographic variables, the scalelets, and the interactions between scalelets and gender was interpreted.

Significant predictors were high-school GPA ($B = .705$, $p < .001$), gender ($B = .196$, $p = .002$), Course-Related Interaction ($B = -.009$, $p = .010$), the interaction between Course Challenge and gender ($B = .010$, $p = .026$), the interaction between Writing and gender ($B = -.012$, $p = .035$), the interaction between Course-Related Interaction and gender ($B = .012$, $p = .003$), and the interaction between Out-of-Class Interaction and gender ($B = -.009$, $p = .032$).

Simple Slopes Analyses. Simple slopes analysis was used to follow up on the statistically significant interactions of the scalelets with gender. For the Course Challenge

scalelet (Figure 1), the simple intercept for males was 2.9952 ($p < .001$), while the simple intercept for females was 3.1914 ($p < .001$). The simple slope for males was -.0037 ($p = .311$), while the simple slope for females was .0064 ($p = .018$). For the Writing scalelet (Figure 2), the simple intercept for males was 2.9952 ($p < .001$), while the simple intercept for females was 3.1914 ($p < .001$). The simple slope for males was .0077 ($p = .110$), while the simple slope for females was -.0040 ($p = .138$). For the Course-Related Interactions with Faculty scalelet (Figure 3), the simple intercept for males was 2.9952 ($p < .001$), while the simple intercept for females was 3.1914 ($p < .001$). The simple slope for males was -.0088 ($p = .010$), while the simple slope for females was .0034 ($p = .133$). For Out-of-Class Interactions (Figure 4), the simple intercept for males was 2.9952 ($p < .001$), while the simple intercept for females was 3.1914 ($p < .001$). The simple slope for males was .0016 ($p = .647$), while the simple slope for females was -.0072 ($p = .002$).

Prediction of Retention Status Using Benchmarks.

Logistic regressions were conducted to examine the relationship between the NSSE subscales and retention to the third semester of college. Prior to the analyses, the following assumptions were assessed: 1) the dependent variable is dichotomous; 2) the independent variables are not multicollinear; 3) independence of errors, and 4) proper model specification. Multicollinearity of independent variables was assessed by looking at the correlations of the independent variables (Tables 1 and 2). Independence of errors was assessed by examining the design of the retention measure and determining that the retention status of one student has minimal or minor impact on the retention status of another. Proper model specification was assessed by including all variables that are theorized to be relevant to the model. The logistic regressions were constructed in a

Figure 1: Course Challenge by Gender Interaction

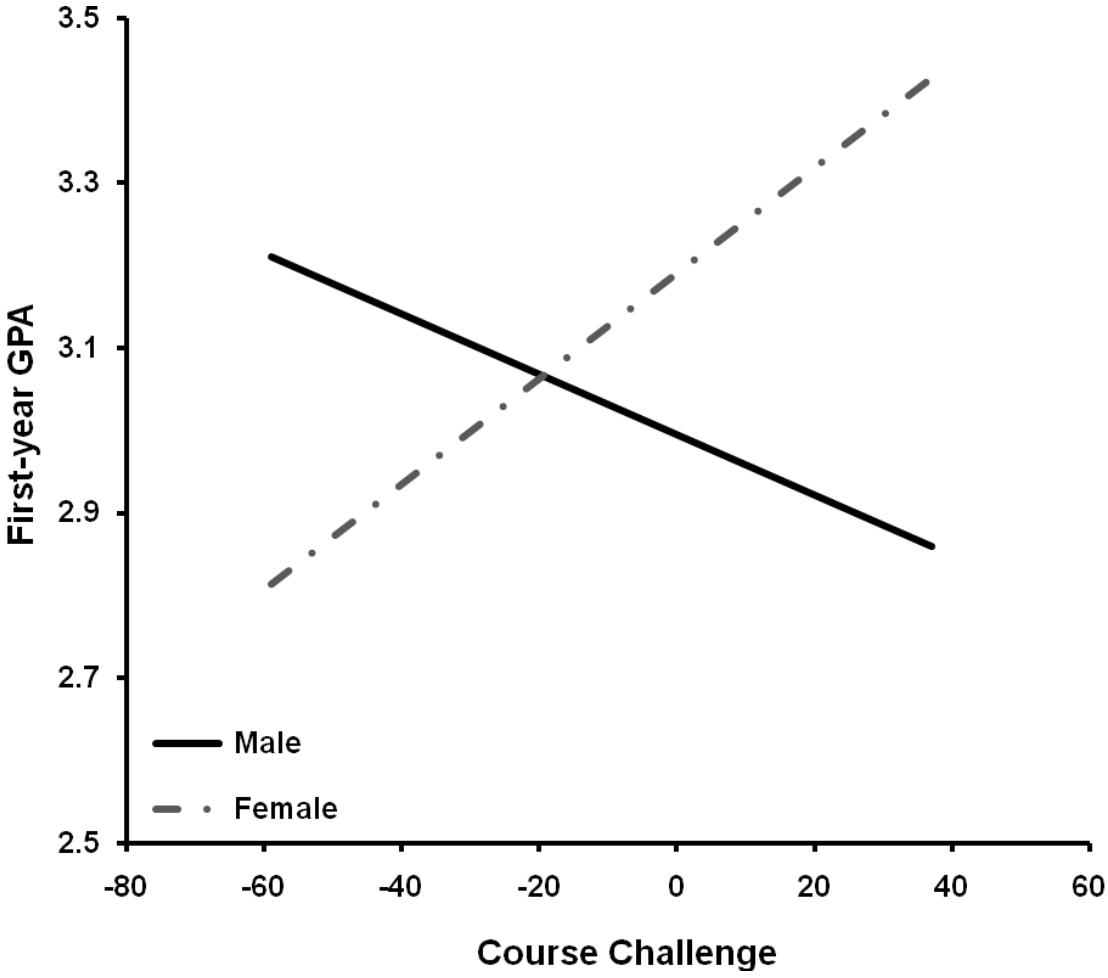


Figure 2: Writing by Gender Interaction

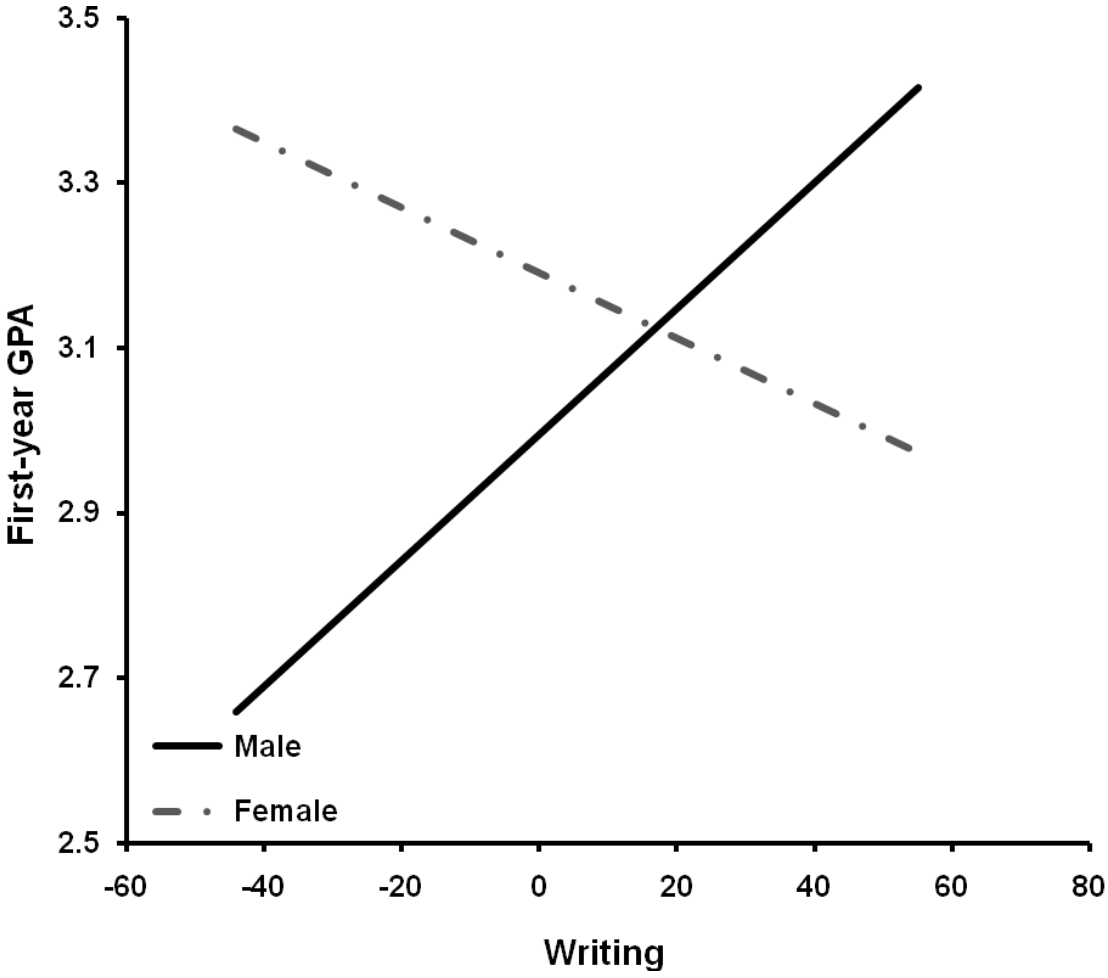


Figure 3: Course-Related Interactions with Faculty by Gender Interaction

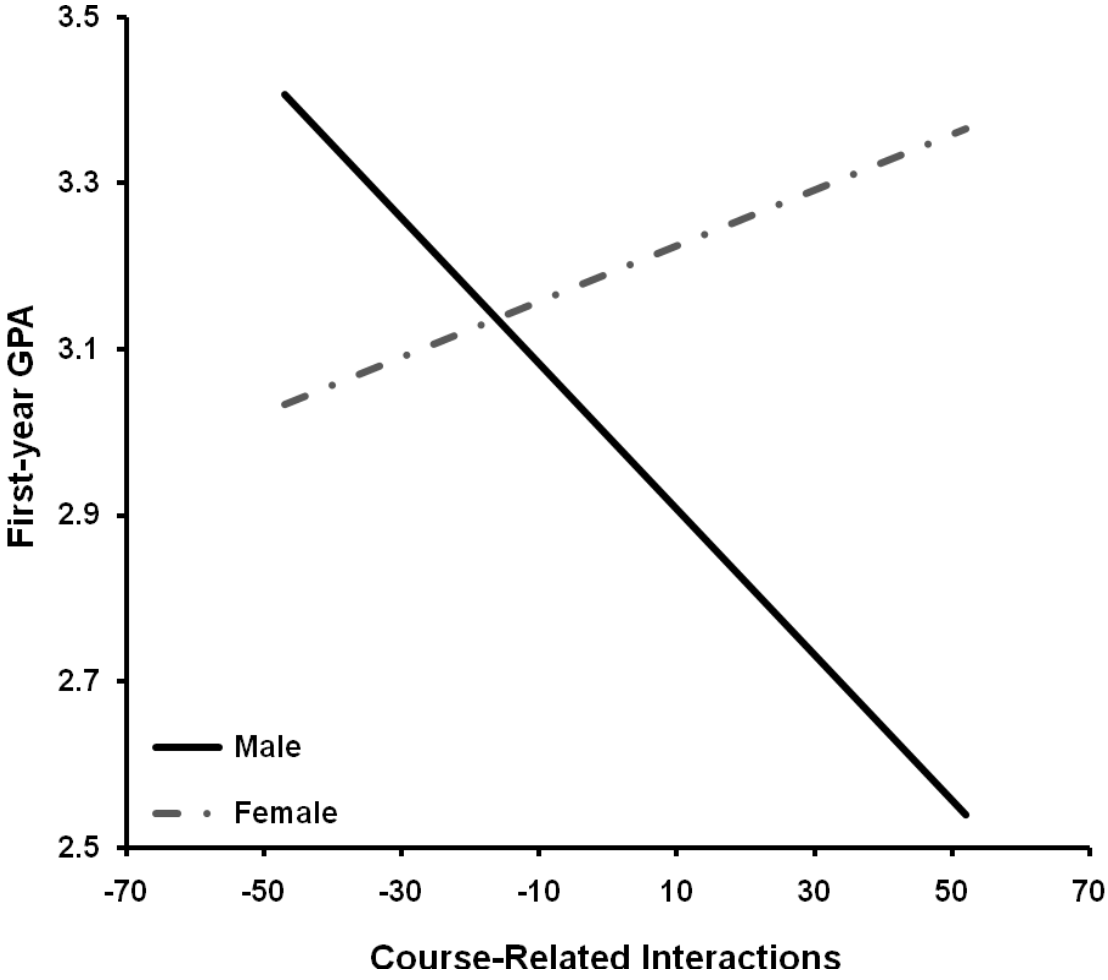
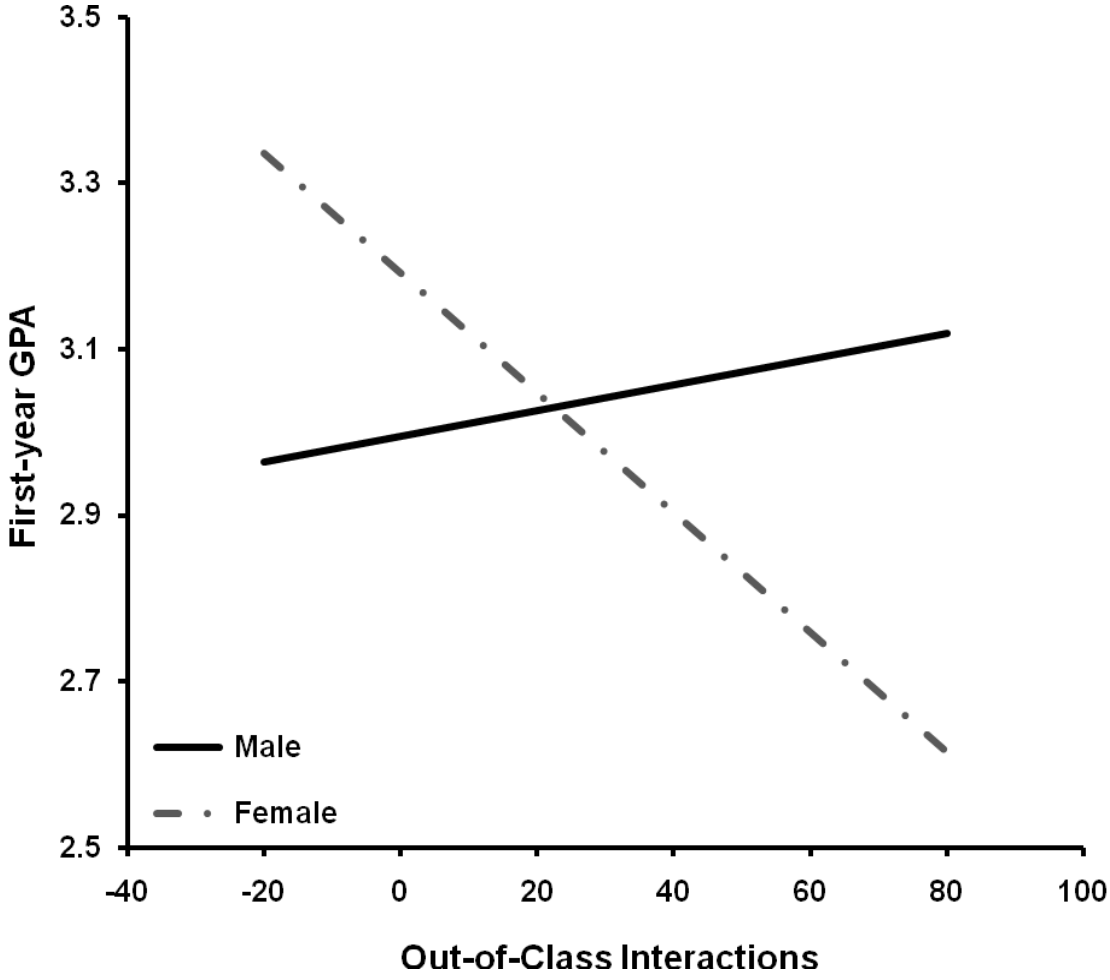


Figure 4: Out-of-Class Interactions with Faculty by Gender Interaction



similar way to the multiple regressions. Variables were entered into the model using a block method of entry, with demographics in block 1, subscales in block 2, and interactions in block 3. Within the ethnicity variable, two of the categories (African American and no response) contained only students who were retained. This meant that there were zero students in each of those two ethnicity category who were also in the not retained category, which prevents the analysis from converging on one result. To prevent this, all students in these two ethnicity categories were not included in the analysis. The benchmark analysis included the remaining 393 students who had complete data, while the scalelet analysis included 371 students.

The first logistic regression was conducted using the benchmarks in blocks 2 and 3. The first model, containing only the first block with the demographic variables, was not statistically significant, X^2 (6, N = 393) = 3.577, p = .734. The second model, containing the first block and the second block with the benchmarks was not statistically significant, X^2 (11, N = 393) = 9.037, p = .619. The third model, containing the first two blocks plus the third block with the interaction variables, was not statistically significant, X^2 (16, N = 393) = 12.304, p = .723.

Prediction of Retention Status Using Scalelets.

The second logistic regression was conducted in the same way as the first logistic regression, using the scalelets in place of the benchmarks. The analysis included the 393 students who had complete data. The first model, using the demographic variables, was not statistically significant, X^2 (6, N = 371) = 3.869, p = .694. The second model, containing the first block and the second block with the scalelets was not statistically significant, X^2 (18, N = 371) = 12.339, p = .829. The third model, containing the first two

blocks plus the third block with interactions, was not statistically significant, X^2 (30, N = 371) = 26.829, p = .632.

Chapter 4

Discussion

The purpose of this study was to investigate the predictive validity of NSSE subscales in relation to student outcomes. Specifically, it investigated how NSSE benchmarks and scalelets relate to first-year GPA and retention to the third semester. It also compared the predictive ability of the benchmarks and scalelets. The relationship between the characteristics gender, race/ethnicity, and high school GPA with first-year GPA and retention was also examined, including the interactions between gender and the subscales.

The first research question was about how the NSSE benchmarks relate to retention and first-year GPA. The NSSE benchmarks were found to account for 2.0% of the variance in first year GPA, after the control variables were accounted for. Although the addition of all the benchmarks significantly increased the variance accounted for, the only benchmark that was found to be related was Student-Faculty Interaction. In addition, the Student-Faculty Interaction Benchmark was negatively related to first-year GPA. This indicates that as interaction with faculty increases, GPA decreases. One explanation for this relationship may be that the students who are having difficulty may be more likely to interact with faculty to get help. However, only one of the six items in the benchmark (see Appendix B), "discussed grades or assignments with an instructor," is relevant to this explanation. Without further information, this relationship is difficult to interpret.

The second research question was about how the scalelets relate to retention and first-year GPA. Pike's scalelets were found to account for 4.2% of variance in first-year GPA, after the control variables were accounted for. The inclusion of the gender/scalelet

interactions accounted for an additional 4.1% of the variance. Since the model containing the interactions between the scalelets and gender significantly increased the variance accounted for, the model containing all variables and interactions was interpreted, using simple slopes analysis to further investigate the gender by scalelet interactions.

The simple slopes analyses for the interactions found mixed results. Four of the scalelets significantly interacted with gender: Course Challenge, Writing, Course-Related Interactions with Faculty, and Out-of-Class Interactions with Faculty. The simple slopes analysis for the Course Challenge scalelet showed a significant positive relationship with first-year GPA for females, while for males Course Challenge did not have a significant relationship. Two of the five items in the scalelet relate to the difficulty of classes and instructor expectations, while the other three relate to the time and effort of students. Interpretation of these results is difficult, as it seems reasonable that students who work harder, putting in more time and effort, would perform better and get better grades. However, it is not intuitive that this relationship would exist only for females and not for males.

The simple slopes analysis for the Writing scalelet found that the relationship with first-year GPA is not significantly different from zero for either males or females, but the intercepts and slopes for males and females are significantly different from each other, with males having a positive slope and females having a negative slope. Three items in this scalelet are about the quantity of writing required, while the other two are about effort and expectations. Therefore, a high score in writing indicates that the student is writing many papers, is writing multiple drafts of those papers, and the assignments required integrating multiple sources. The results indicate that students who score high in

Writing perform well if they are male, but perform poorly if they are female. Again, interpretation is difficult, as this is in contrast to the gender by Course Challenge interaction, which found that females perform better with more challenge.

The simple slopes analysis for the Course-Related Interactions with Faculty scalelet found that there is a significant negative relationship with first-year GPA for males, but no relationship for females. This means that for males, as Course-Related Interactions go up, GPAs go down. The items in this scalelet are related to discussing grades, assignments, readings and ideas from class, and receiving feedback from the instructor. This result is similar to that of the benchmarks analysis, with Student-Faculty Interaction. However, in this case, only males have a significant relationship.

For the Out-of-Class Interactions with Faculty scalelet, the simple slopes analysis found that there is a significant negative relationship with first-year GPA for females, but not for males. This means that for females, as Out-of-Class Interactions with Faculty go up, GPAs go down. The items in the scalelet are related to discussing career plans or working with faculty on outside research or other projects unrelated to class. As freshmen are unlikely to be participating in these activities, most of the students surveyed had not spent any time on these activities, which may mean that the results are driven by a few students.

An interesting aspect of these results is the similar, yet opposite relationships with faculty for males and females. For males, interactions with faculty that are related to class are related to lower grades, while for females, it is the interactions with faculty that are unrelated to class that lead to lower grade. However, in both cases, it is difficult to interpret the results without more information, and more theory about why these

relationships would exist.

The results of this study show small relationships, both negative and positive, between NSSE subscales and first-year GPA. These results are similar to those found in Gordon, et al (2008), who also found a mix of positive and negative predictors of first-year GPA. However, the significant benchmarks and scalelets are not the same for any of the analyses conducted. While most of the negative predictors in this study were related to the interaction between students and faculty, Gordon, et al, found negative relationships with first-year GPA and the Enriching Educational Experiences benchmark, as well as the Information Technology scalelet and the Diversity scalelet, which share many items with the Enriching Educational Experiences benchmark.

Other studies that examined similar relationships found only positive predictors of outcomes (Pike, et al, 2006; Carini, et al, 2006). These studies also found much stronger relationships between NSSE subscales and student outcomes. However, these studies differ from the present study in many ways. The Pike, et al, 2006, study used institution, rather than student, as the unit of analysis. It also used the General Education and Practical Skills sections of the NSSE as outcome variables. The Carini, et al, 2006, study uses correlations instead of regressions to test the relationships between NSSE subscales and student outcomes. These differences may be why the studies have such different results.

The first two research questions also address the relationships between the NSSE subscales and retention. The NSSE benchmarks were found to be unrelated to retention to the third semester, with none of the benchmarks significant predictors, and the model was non-significant. The same result was found with Pike's scalelets. However, Gordon, et al

(2008) found that their control variable, the Supportive Campus Environment benchmark, the Out-of-Class Interaction scalelet, the varied experiences scalelet and the Interpersonal environment scalelet all predicted retention. It seems likely that the study's measures were not ideal to find these relationships. The survey was administered in the students' second semester at school, and retention was measured at the third semester. As the University of New Mexico has many programs for first-year students that are designed to help them and keep them enrolled in school, it is likely that the students who are struggling and may eventually drop out are still retained in the third semester. Additionally, the first-year students who did not return for the third semester may have already made the decision in their second semester, and so would be less likely to choose to participate in the survey, as they would have no stake in the results or sense of obligation to participate. A final problem in the retention measure is that the students who did not return for their third semester may have done so for multiple reasons, including transferring to a different school, which may be entirely unrelated to their level of engagement, or demographic characteristics.

The third research question addresses how the scalelets and benchmarks compare in their ability to predict educational outcomes. Based on these analyses, Pike's scalelets account for slightly more variance in first-year GPA than the NSSE benchmarks. The final model in the benchmark analysis accounted for 32.2% of the variance in first-year GPA, with 2.0% of that attributed to the addition of the benchmarks to the model. The final model in the scalelet analysis accounted for 38.8% of the variance in first-year GPA, with 8.2% of that attributed to the addition of the scalelets and the interactions between the scalelets and gender. Neither the scalelets nor the benchmarks were significant

predictors of retention, so comparison was not relevant.

Research questions 4, 5, and 6 ask how high school GPA, race/ethnicity and gender relate to first-year GPA, while the final research question asks how gender interacts with the subscales. High school GPA was a significant predictor for first-year GPA for both the scalelet and the benchmark analyses, showing a direct relationship between high school GPA and college GPA. However, high school GPA was not significantly predictive of retention. Race and ethnicity were found to have no relationship with either first-year GPA or retention, in that none of the six race/ethnicity variables were significant predictors in any of the analyses. Gender was a significant predictor of first-year GPA, but not retention, with females scoring higher than males. In addition, while the benchmarks did not interact with gender in predicting GPA, the scalelets analysis found four significant interactions, as described above.

Strengths and Limitations

As with all studies, there are many strengths and limitations to this study. The primary strength of the study is the sample of students who responded to the survey was relatively large. A large sample size increases our ability to find statistically significant effects. It also enables the inclusion of many variables in the regression analysis, such as the interaction variables.

However, a limitation of so many variables is that it increases the risk of type I error. Each analysis involved three models, each of which had seven to 33 coefficients. The first analysis, using benchmarks and GPA, produced 42 coefficients, each of which was statistically tested for a difference from zero. This high number of statistical tests means that it is probable that at least one test will show a statistical difference that is due

purely to chance. The same is true for all four analyses. Therefore, it is likely that at least some of these significant results are due to chance fluctuations of the data, rather than a meaningful relationship between variables.

Another limitation of the study is that the students who responded to the survey differed from those who did not respond in several characteristics. The respondents had higher high school GPAs, higher first-year GPAs, were more likely to be female, and were more likely to return to UNM for a third semester. This means that the sample was not representative of the general population. This is in contrast to the studies by IUCPR (NSSE, 2008), which found only trivial differences in characteristics of respondents and non-respondents. As a result of these differences, it is unreasonable to infer any results to the general population, even the specific population of first-year students at the University of New Mexico. However, one use of the NSSE is to compare results with other schools. It is likely that most schools, especially those with which UNM has the most in common, face the same problem with their sample. Therefore, it may be reasonable to compare the first-year students at UNM who elected to respond to the NSSE with first-year students at a different university who elected to respond to the NSSE, and make a meaningful comparison.

The limitations of survey research in general are apparent in this study. Many of the objections to survey research, as detailed by Porter (2009), were found to be reasonable here. Although the NSSE reports average response rates of about 40% (Kuh, 2001), we found response rates of about 23%. Despite efforts to make the survey more attractive and easier to complete, our response rate was still quite low. Kuh (2001) describes the main concern of low response rates to be a bias in the characteristics of

those who respond compared to those who do not (Chen, et al, 2009). Although studies by IUCPR (2008) have found minor differences in characteristics and behaviors of respondents and non-respondents, we found significant differences between our groups. In addition, the process of making the survey easier and more attractive resulted in many of the questions in the survey being unnecessarily vague, potentially resulting in different understandings of the questions between different students. This may be why some subscales had such low reliability scores.

Another limitation is that the NSSE is intended to be administered to first-year students and graduating seniors. Therefore, the freshmen sample may have very little variation in their responses to some questions about experiences. For example, all three of the questions in the Out-of-Class Interactions with Faculty scalelet may not be relevant to first-year students. They may not be ready to discuss career plans, and are probably focused on coursework, not yet working on outside projects or research. This abnormal distribution may mean that the results are biased by the responses of the few students who reported that they have done the activities in question.

A final issue is that this analysis is based on only one year of data, from one class of students. This means that there is limited data for many of the variables. The NSSE data comes from just one year of students, measured once. Data for cumulative GPA is based on only two semesters, so one or two unusual classes may be biasing the scores. Retention may be most affected, as three semesters may not be long enough for students to drop out of school, especially with the programs at UNM that are aimed at retaining students beyond the first year.

Implications for Future Research

The main implication of these findings is that more research needs to be done. The first step should be similar analyses using data from previous years to confirm the findings in this study. A follow up to this study using GPA and enrollment data from later semesters would also help confirm the results found here.

However, even if these results were confirmed using similar data from other years, they still could not be inferred to the general population. Future administrations of the NSSE should attempt to increase the response rate in order to improve the representativeness of the sample. They should also attempt to improve incentives and focus on recruiting underrepresented students.

Conclusions

This study evaluated the predictive validity of the NSSE benchmarks and scalelets. Although the benchmarks were able to predict a small amount of variance in first-year GPA, and the scalelets were able to predict a small amount more variance than the benchmarks, neither were able to predict a significant amount of variance in retention. The many limitations of this study mean that the interpretation of these results must be minimal. Any significant relationship found in the analyses is not necessarily meaningful. Based on the evidence found here, there is little evidence for predictive validity.

Appendix A: National Survey of Student Engagement

1. Academic and Intellectual Experiences

In your experience at your institution during the current school year, about how often have you done each of the following?

1=Never, 2=Sometimes, 3=Often, 4=Very often

a. Asked questions in class or contributed to class discussions
b. Made a class presentation
c. Prepared two or more drafts of a paper or assignment before turning it in
d. Worked on a paper or project that required integrating ideas or information from various sources
e. Included diverse perspectives (different races, religions, genders, political beliefs, etc.) in class discussions or writing assignments
f. Come to class without completing readings or assignments
g. Worked with other students on projects **during class**
h. Worked with classmates **outside of class** to prepare class assignments
i. Put together ideas or concepts from different courses when completing assignments or during class discussions
j. Tutored or taught other students (paid or voluntary)
k. Participated in a community-based project (e.g. service learning) as part of a regular course
l. Used an electronic medium (listserv, chat group, Internet, instant messaging, etc.) to discuss or complete an assignment
m. Used e-mail to communicate with an instructor
n. Discussed grades or assignments with an instructor
o. Talked about career plans with a faculty member or advisor
p. Discussed ideas from your readings or classes with faculty members outside of class
q. Received prompt written or oral feedback from faculty on your academic performance
r. Worked harder than you thought you could to meet an instructor's standards or expectations
s. Worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.)
t. Discussed ideas from your readings or classes with others outside of class (students, family members, co-workers, etc.)
u. Had serious conversations with students of a different race or ethnicity than your own
v. Had serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values

2. Mental Activities

During the current school year, how much has your coursework emphasized the following mental activities?

1=Very little, 2=Some, 3=Quite a bit, 4=Very much

a. **Memorizing** facts, ideas, or methods from your courses and readings so you can repeat them in pretty much the same form
b. **Analyzing** the basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components
c. **Synthesizing** and organizing ideas, information, or experiences into new, more complex interpretations and relationships
d. **Making judgments** about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusions
e. **Applying** theories or concepts to practical problems or in new situations

3. Reading and Writing

During the current school year, about how much reading and writing have you done?

1=None, 2=1-4, 3=5-10, 5=More than 20

a. Number of assigned textbooks, books, or book-length packs of course readings
b. Number of books read on your own (not assigned) for personal enjoyment or academic enrichment
c. Number of written papers or reports of **20 pages or more**
d. Number of written papers or reports **between 5 and 19 pages**
e. Number of written papers or reports of **fewer than 5 pages**

4. Problem Sets

In a typical week, how many homework problem sets do you complete?

1=None, 2=1-2, 3=3-4, 4=5-6, 5=More than 6

a. Number of problem sets that take you **more** than an hour to complete
b. Number of problem sets that take you **less** than an hour to complete

5. Examinations

1=Very little to 7=Very much

Select the circle that best represents the extent to which your examinations during the current school year challenged you to do your best work.

6. Additional Collegiate Experiences

During the current school year, about how often have you done each of the following?

1=Never, 2=Sometimes, 3=Often, 4=Very often

a. Attended an art exhibit, play, dance, music, theatre or other performance
b. Exercised or participated in physical fitness activities
c. Participated in activities to enhance your spirituality (worship, meditation, prayer, etc.)
d. Examined the strengths and weaknesses of your own views on a topic or issue
e. Tried to better understand someone else's views by imagining how an issue looks from his or her perspective
f. Learned something that changed the way you understand an issue or concept

7. Enriching Educational Experiences

Which of the following have you done or do you plan to do before you graduate from your institution?

0=Have not decided, Do not plan to do, Plan to do; 1=Done.

a. Practicum, internship, field experience, co-op experience, or clinical assignment
b. Community service or volunteer work
c. Participate in a learning community or some other formal program where groups of students take two or more classes together
d. Work on a research project with a faculty member outside of course or program requirements
e. Foreign language coursework
f. Study abroad
g. Independent study or self-designed major
h. Culminating senior experience (capstone course, senior project or thesis, comprehensive exam, etc.)

8. Quality of Relationships

Select the circle that best represents the quality of your relationships with people at your institution.

1=Unfriendly, Unsupportive, Sense of alienation to 7=Friendly, Supportive, Sense of belonging

a. Relationships with **other students**
   1=Unavailable, Unhelpful, Unsympathetic to 7=Available, Helpful, Sympathetic
b. Relationships with **faculty members**
   1=Unhelpful, Inconsiderate, Rigid to 7=Helpful, Considerate, Flexible
c. Relationships with **administrative personnel and offices**

9. Time Usage

About how many hours do you spend in a typical 7-day week doing each of the following?

1=0 hrs/wk, 2=1-5 hrs/wk, 3=6-10 hrs/wk, 4=11-15 hrs/wk, 5=16-20 hrs/wk, 6=21-25 hrs/wk, 7=26-30 hrs/wk, 8=More than 30 hrs/wk

a. Preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities)
b. Working for pay **on campus**
c. Working for pay **off campus**
d. Participating in co-curricular activities (organizations, campus publications, student government, fraternity or sorority, intercollegiate or intramural sports, etc.)
e. Relaxing and socializing (watching TV, partying, etc.)
f. Providing care for dependents living with you (parents, children, spouse, etc.)
g. Commuting to class (driving, walking, etc.)

10. Institutional Environment

To what extent does your institution emphasize each of the following?

1=Very little, 2=Some, 3=Quite a bit, 4=Very much

a. Spending significant amounts of time studying and on academic work
b. Providing the support you need to help you succeed academically
c. Encouraging contact among students from different economic, social, and racial or ethnic backgrounds
d. Helping you cope with your non-academic responsibilities (work, family, etc.)
e. Providing the support you need to thrive socially
f. Attending campus events and activities (special speakers, cultural performances, athletic events, etc.)
g. Using computers in academic work

11. Educational and Personal Growth

To what extent has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas?

1=Very little, 2=Some, 3=Quite a bit, 4=Very much

a. Acquiring a broad general education
b. Acquiring job or work-related knowledge and skills
c. Writing clearly and effectively
d. Speaking clearly and effectively
e. Thinking critically and analytically
f. Analyzing quantitative problems
g. Using computing and information technology
h. Working effectively with others
i. Voting in local, state, or national elections
j. Learning effectively on your own
k. Understanding yourself

l. Understanding people of other racial and ethnic backgrounds
m. Solving complex real-world problems
n. Developing a personal code of values and ethics
o. Contributing to the welfare of your community
p. Developing a deepened sense of spirituality

12. Academic Advising

1=Poor, 2=Fair, 3=Good, 4=Excellent

Overall, how would you evaluate the quality of academic advising you have received at your institution?

13. Satisfaction

1=Poor, 2=Fair, 3=Good, 4=Excellent

How would you evaluate your entire educational experience at this institution?

14. *1=Definitely no, 2=Probably no, 3=Probably yes, 4=Definitely yes*

If you could start over again, would you go to the **same institution** you are now attending?

Appendix B: NSSE Benchmark and Scalelet Items.

Level of Academic Challenge (LAC) Benchmark

Number of assigned textbooks, books, or book-length packs of course readings

Course Challenge (CC) Scalelet

Worked harder than you thought you could to meet an instructor's standards or expectations

Hours per week spent preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities)

Spending significant amounts of time studying and on academic work

Come to class without completing readings or assignments*

Select the circle that best represents the extent to which your examinations during the current school year challenged you to do your best work.*

Higher Order Thinking Skills (HOTS) Scalelet

Analyzing the basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components

Synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships

Making judgments about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusions

Applying theories or concepts to practical problems or in new situations

Memorizing facts, ideas, or methods from your courses and readings so you can repeat them in pretty much the same form*

Writing (Wr)Scalelet

Number of written papers or reports of 20 pages or more

Number of written papers or reports between 5 and 19 pages

Number of written papers or reports of fewer than 5 pages

Prepared two or more drafts of a paper or assignment before turning it in*

Worked on a paper or project that required integrating ideas or information from various sources*

Active and Collaborative Learning (ACL) Benchmark

Active Learning (AL) Scalelet

Asked questions in class or contributed to class discussions

Made a class presentation

Participated in a community-based project (e.g. service learning) as part of a regular course

Collaborative Learning (CL) Scalelet

Worked with other students on projects during class

Worked with classmates outside of class to prepare class assignments

Tutored or taught other students (paid or voluntary)

Discussed ideas from your readings or classes with others outside of class (students, family members, co-workers, etc.)

Student-Faculty Interaction (SFI) Benchmark

Course-Related Interactions with Faculty Scalelet

Discussed grades or assignments with an instructor

Discussed ideas from your readings or classes with faculty members outside of class

Received prompt written or oral feedback from faculty on your academic performance

Out-of-Class Interactions with Faculty Scalelet

Talked about career plans with a faculty member or advisor

Worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.)

Work on a research project with a faculty member outside of course or program requirements

Enriching Educational Experiences (EEE) Benchmark

Diversity (Div) Scalelet

Had serious conversations with students of a different race or ethnicity than your own

Had serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values

Encouraging contact among students from different economic, social, and racial or ethnic backgrounds

Information Technology (IT) Scalelet

Used an electronic medium (listserv, chat group, Internet, instant messaging, etc.) to discuss or complete an assignment

Used e-mail to communicate with an instructor*

Using computers in academic work*

Varied Educational Experiences (VE) Scalelet

Practicum, internship, field experience, co-op experience, or clinical assignment

Community service or volunteer work

Participate in a learning community or some other formal program where groups of students take two or more classes together

Foreign language coursework

Study abroad

Independent study or self-designed major

Culminating senior experience (capstone course, senior project or thesis, comprehensive exam, etc.)

Participating in co-curricular activities (organizations, campus publications, student government, fraternity or sorority, intercollegiate or intramural sports, etc.)

Attending campus events and activities (special speakers, cultural performances, athletic events, etc.)*

Supportive Campus Environment (SCE) Benchmark

Interpersonal Environment (IE) Scalelet

Relationships with other students

Relationships with faculty members

Relationships with administrative personnel and offices

Support for Student Success (SSS) Scalelet

Providing the support you need to help you succeed academically

Helping you cope with your non-academic responsibilities (work, family, etc.)

Providing the support you need to thrive socially

* Indicates item included in scalelet but not in benchmark.

Appendix C: Regression Results

Table 6: Regression coefficients for Benchmarks and First-year GPA.

Explanatory Variable	Model 1	Model 2	Model 3
	B (β)	**B (β)**	**B (β)**
Intercept	3.038	3.035	3.029
Demographics			
High School GPA	.739 (.520)***	.702 (.494)***	.704 (.496)***
Gender	.160 (.112)**	.162 (.113)**	.163 (.114)**
Ethnicity (ref=White, non-Hispanic)			
African American/Black	.077 (.023)	.086 (.025)	.106 (.031)
American Indian	.030 (.010)	.023 (.008)	.040 (.013)
Asian/Pacific Islander	-.211 (-.058)	-.254 (-.070)	-.247 (-.068)
Hispanic	-.026 (-.019)	-.014 (-.011)	-.009 (-.007)
International	-.056 (-.011)	-.055 (-.011)	-.066 (-.013)
No response	.197 (.039)	.109 (.022)	.143 (.028)
Benchmarks			
Level of Academic Challenge		.002 (.038)	.003 (.063)
Active and Collaborative Learning		.003 (.082)	.002 (.057)
Student-Faculty Interaction		-.005 (-.128)*	-.006 (-.169)
Enriching Educational Experiences		.005 (.097)	.009 (.178)
Supportive Campus Environment		.000 (-.004)	-.003 (-.090)
Interactions			
Benchmark*Gender			
Level of Academic Challenge			-.002 (-.029)
Active and Collaborative Learning			.001 (.020)
Student-Faculty Interaction			.002 (.057)
Enriching Educational Experiences			-.006 (-.099)
Supportive Campus Environment			.005 (.108)
R^2	.303***	.323***	.328***
R^2 Change		.020*	.005

Notes: B, Raw Coefficients; β, Standardized Coefficients. * p < .05, ** p < .01, *** p < .001

Table 7: Regression Coefficients for Scalelets and First-year GPA

Explanatory Variable	Model 1	Model 2	Model 3
	B (β)	**B (β)**	**B (β)**
Intercept	3.030	3.019	2.995
Demographics			
High School GPA	.722 (.517)***	.686 (.492)***	.705 (.505)***
Gender	.178 (.125)**	.185 (.130)**	.196 (.137)***
Ethnicity (ref=White, non-Hispanic)			
African American/Black	-.035 (-.010)	-.009 (-.002)	.047 (.013)
American Indian	.083 (.028)	.070 (.024)	.114 (.039)
Asian/Pacific Islander	-.213 (-.061)	-.202 (-.058)	-.176 (-.050)
Hispanic	-.024 (-.018)	-.009 (-.007)	-.009 (-.007)
International	-.062 (-.013)	-.073 (-.015)	-.076 (-.015)
No response	.234 (-.044)	.180 (.034)	.234 (.044)
Scalelets			
Course Challenge		.002 (.057)	-.004 (-.084)
Writing		-.001 (-.016)	.008 (.166)
Active Learning Experiences		.003 (.100)	.001 (.042)
Collaborative Learning Experiences		.001 (.036)	.001 (.031)
Course-Related Interactions		-.001 (-.023)	-.009 (-.284)*
Out-of-Class Interactions		-.004 (-.126)*	.002 (.048)
Use of Information Technology		.000 (-.010)	.003 (.078)
Emphasis on Diversity		.001 (.038)	.002 (.058)
Varied Educational Experiences		.005 (.106)*	.006 (.125)
Support for Student Success		-.003 (-.117)*	-.002 (-.078)
Interpersonal Environment		.003 (.086)	.000 (-.013)
Higher Order Thinking Skills		-.001 (-.016)	.001 (-.026)
Interactions			
Scalelet*Gender			
Course Challenge			.010 (.182)*
Writing			-.012 (-.216)*
Active Learning Experiences			.003 (.076)
Collaborative Learning Experiences			.000 (-.001)
Course-Related Interactions			.012 (.331)**
Out-of-Class Interactions			-.009 (-.229)*
Use of Information Technology			-.003 (-.084)
Emphasis on Diversity			-.001 (-.038)
Varied Educational Experiences			-.001 (-.026)
Support for Student Success			-.001 (-.027)
Interpersonal Environment			.005 (.118)
Higher Order Thinking Skills			-.003 (-.066)
R^2	.306***	.348***	.388***
R^2 Change		.042*	.041*

Notes: B, Raw Coefficients; β, Standardized Coefficients. * p < .05, ** p < .01, *** p < .001

Table 8: Regression Coefficients for Benchmarks and Retention.

Explanatory Variable	Model 1	Model 2	Model 3
	B Exp(β)	**B Exp(β)**	**B Exp(β)**
Intercept	2.569	2.620	2.845
Demographics			
High School GPA	.173 (1.189)	.176 (1.193)	.227 (1.255)
Gender	-.302 (.739)	-.328 (.720)	-.562 (.570)
Ethnicity (ref=White, non-Hispanic)			
American Indian	.673 (1.960)	.656 (1.927)	.686 (1.986)
Asian/Pacific Islander	-.986 (.373)	-.945 (.389)	-.999 (.368)
Hispanic	-.296 (.744)	-.253 (.776)	-.274 (.761)
International	-.543 (.581)	-.510 (.600)	-.621 (.537)
Benchmarks			
Level of Academic Challenge		-.029 (.972)	-.061 (.941)
Active and Collaborative Learning		.017 (1.017)	.028 (1.029)
Student-Faculty Interaction		-.005 (.995)	.018 (1.018)
Enriching Educational Experiences		.018 (1.019)	.040 (1.041)
Supportive Campus Environment		.011 (1.011)	-.004 (.996)
Interactions			
Benchmark*Gender			
Level of Academic Challenge			.046 (1.047)
Active and Collaborative Learning			-018 (.982)
Student-Faculty Interaction			-.027 (.973)
Enriching Educational Experiences			-.027 (.973)
Supportive Campus Environment			.020 (1.021)
X^2	4.485(p = .611)	9.349(p = .590)	12.509(p = .708)
X^2 Change		4.865(p = .433)	3.160(p = .675)

Notes: B, Raw Coefficients; Exp(β), Odds Ratios. * p < .05, ** p < .01, *** p < .001

Table 9: Regression Coefficients for Scalelets and Retention.

Explanatory Variable	Model 1	Model 2	Model 3
	B Exp(β)	**B Exp(β)**	**B Exp(β)**
Intercept	2.578	2.660	3.721
Demographics			
High School GPA	.159 (1.173)	.038 (1.039)	.147 (1.159)
Gender	-.286 (.751)	-.277 (.758)	-1.252 (.286)
Ethnicity (ref=White, non-Hispanic)			
American Indian	.595 (1.813)	.441 (1.555)	.228 (1.256)
Asian/Pacific Islander	-1.006 (.366)	-.904 (.405)	-1.226 (.293)
Hispanic	-.409 (.664)	-.415 (.661)	-.491 (.612)
International	-.566 (.568)	-.356 (.701)	-.635 (.530)
Scalelets			
Course Challenge		.002 (1.002)	-.008 (.992)
Writing		-.009 (.991)	-.003 (.997)
Active Learning Experiences		.009 (1.009)	.006 (1.006)
Collaborative Learning Experiences		.010 (1.010)	.015 (1.015)
Course-Related Interactions		-.013 (.987)	.037 (1.038)
Out-of-Class Interactions		.009 (1.009)	-.018 (.982)
Use of Information Technology		.007 (1.007)	.050 (1.051)
Emphasis on Diversity		.006 (1.006)	.005 (1.005)
Varied Educational Experiences		.013 (1.013)	.044 (1.045)
Support for Student Success		-.002 (.998)	-.012 (.988)
Interpersonal Environment		.008 (1.008)	.012 (1.012)
Higher Order Thinking Skills		-.020 (.980)	-.080 (.923)*
Interactions			
Scalelet*Gender			
Course Challenge			.020 (1.020)
Writing			-.009 (.991)
Active Learning Experiences			-.003 (.997)
Collaborative Learning Experiences			-.010 (.990)
Course-Related Interactions			-.062 (.940)
Out-of-Class Interactions			.037 (1.038)
Use of Information Technology			-.055 (.947)
Emphasis on Diversity			.002 (1.002)
Varied Educational Experiences			-.034 (.966)
Support for Student Success			.013 (1.013)
Interpersonal Environment			-.006 (.994)
Higher Order Thinking Skills			.076 (1.079)
X^2	3.869(p = .694)	12.339(p = .829)	26.829(p = .632)
X^2 Change		8.470(p = .747)	14.490(p = .271)

Notes: B, Raw Coefficients; β, Standardized Coefficients. * p < .05, ** p < .01, *** p < .001

References

Aaker, D.A., Kumar, V., & Day, G.S. (1998). *Marketing research* (6 ed.). New York: Wiley.

American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Astin, A.W. (1993). What matters in college? *Liberal Education*, *79*(4), database: Education Research Complete.

Bradburn, N.M., & Sudman, S. (1988). *Polls and surveys: Understanding what they tell us*. San Francisco: Jossey-Bass.

Brandt, R. M. (1958). The accuracy of self estimates. *Genetic Psychology Monographs, 58*, 55-99.

Carini, R.M., Kuh, G.D., & Klein, S.P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, *47*(1), 1-32.

Chen, P.D., Gonyea, R.M., Sarraf, S.A., BrckaLorenz, A., Korkmaz, A., Lambert, A.D., Shoup, R., & Williams, J.M. (2009). Analyzing and interpreting NSSE data. *New Directions in Institutional Research*, *2009*(141), 35-54.

Chickering, A.W., and Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *39*(7), 3-7.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavior Sciences*. Lawrence Erlbaum Associates.

Converse, J.M., & Presser, S. (1989). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psykometrika*, *16*, 297–333.

DeNisi, A.S., & Shaw, J.B. (1977). Investigation of the uses of self-reports of abilities. *Journal of Applied Psychology*, *62*, 641-644.

Ethington, C.A., Horn, R.A. (2007). An Examination of Pace's Model of Student Development and College Impress. *Community College Journal of Research & Practice*, *31*(3), 183-198.

Gordon, J., Ludlum, J., & Hoey, J.J. (2008). Validating NSSE against student outcomes: Are they related? *Research in Higher Education*, *49*(1), 19-39.

Hansford, B.C., & Hattie, J.A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, *52*(1), 123-142.

Indiana University Center for Postsecondary Research (2009). *Reliability*. http://nsse.iub.edu/html/PsychometricPortfolio_Reliability.cfm#internal_consistency_of_nsse_benchmarks. Accessed November 11, 2009.

Kuh, G.D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change*, *33*(3), 10-18.

Kuh, G.D. (2004). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties*. Bloomington, IN: Indiana University, Center for Postsecondary Research.

Kuh, G.D. (2003). What we're learning about student engagement from NSSE. *Change*, *35*(2), 24-32.

Kuh, G.D. (2009). The national survey of student engagement: Conceptual and empirical

foundations. *New Directions in Institutional Research*, *2009*(141), 5-20.

Kuh, G.D., Cruce, T.M., Shoup, R., Kinzie, J., & Gonyea, R.M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, *79*(5), 540-563.

Kuh, G.D., Hayek, J.C., Carini, R.M., Ouimet, J.A., Gonyea, R.M., and Kennedy, J. (2001). *NSSE Technical and Norms Report*. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Kuh, G.D., Kinzie, J., Cruce, T., Shoup, R., & Gonyea, R.M. (2007). *Connecting the dots: Multi-faceted analyses of the relationships between student engagement results from the NSSE, and the institutional practices and conditions that foster student success*. Bloomington, IN: Indiana University, Center for Postsecondary Research.

Kuh, G.D., Pace, C.R. and Vesper, N. (1997). The development of process indicators to estimate student gains associated with good practices in undergraduate education. *Research in Higher Education, 38*(4), 435-454.

Kuh, G.D., Schuh, J.H., Whitt, E.J., and Associates (1991). *Involving Colleges: Successful Approaches to Fostering Student Learning and Personal Development Outside the Classroom*. San Francisco: Jossey-Bass, 1991.

Laing, J., Swayer, R, & Noble, J. (1987). Accuracy of self-reported activities and accomplishments of college-bound seniors. *Journal of College Student Development*, *29*(6), 362-368.

LaNasa, S.M., Cabrera, A.F., and Trangsrud, H. (2009). The construct validity of student engagement: A confirmatory factor analysis approach. *Research in Higher*

Education, *50*(4), 315-332.

Lowman, R.L., & Williams, R.E. (1987). Validity of self-ratings of abilities and competencies. *Journal of Vocational Behavior*, *31*(1), 1-13.

Ouimet, J.A., Bunnage, J.C., Carini, R.M., Kuh, G.D., & Kennedy, J. (2004). Using focus groups, expert advice, and cognitive interviews to establish the validity of a college student survey. *Research in Higher Education*, *45*(3), 233-250.

Pace, C.R. (1980). Measuring the quality of student effort. *Current Issues in Higher Education*, *2*(1), 10-16.

Pace, C.R. (1985). *The credibility of student self-reports*. Los Angeles: University of California, The Center for the Study of Evaluation, Graduate School of Education.

Pampel, F.C. (2000). *Logistic Regression: A Primer*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-132. Thousand Oaks, CA: Sage.

Pascarella, E.T., and Terenzini, P.T. (2005). *How College Affects Students: A Third Decade of Research*. San Francisco: Jossey-Bass, 2005.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Holt, Rinehart and Winston.

Pike, G.R. (1995). The relationships between self reports of college experiences and achievement test scores. *Research in Higher Education*, *36*(1), 1-22.

Pike, G.R. (2004). Measuring quality: A comparison of *U.S. News* rankings and NSSE benchmarks. *Research in Higher Education, 45*(2), 193-208.

Pike, G.B. (2006). The convergent and discriminant validity of NSSE scalelet scores. *Journal of College Student Development*, *47*(5), 550-563.

Porter, S.R. (2006). Institutional structures and student engagement. *Research in Higher Education*, *47*(5), 521-558.

Porter, S.R. (2009). Do college student surveys have any validity? Paper presented at the 2009 meeting of the Association for the Study of Higher Education.

Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press.

Wentland, E.J. & Smith, K.W. (1993). *Survey responses: An evaluation of their validity*. New York: Academic Press.