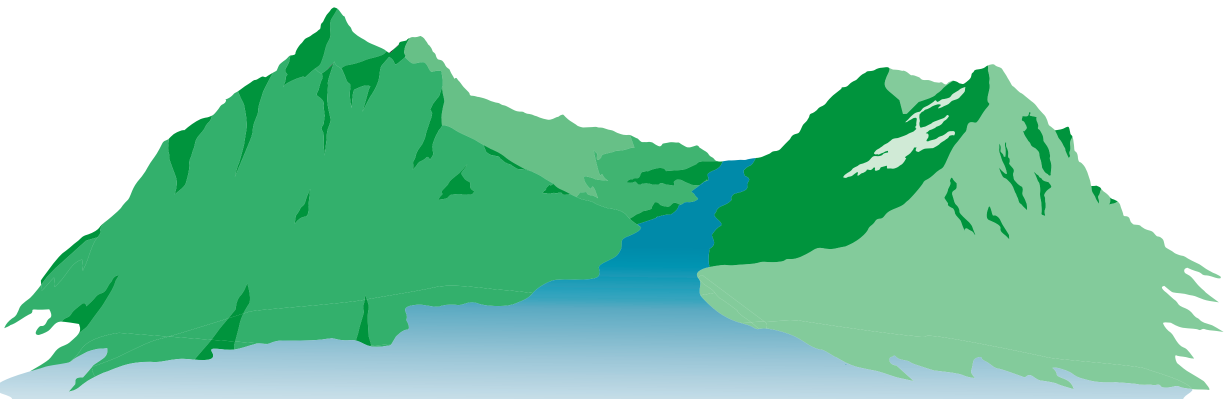
42nd NEAIR Annual Conference Proceedings



INSTITUTIONAL RESEARCH

WHEN THE ONLY CONSTANT IS CHANGE

OCTOBER 31 – NOVEMBER 3, 2015

Sheraton Burlington Hotel & Conference Center
Burlington, Vermont

CONFERENCE TEAM

President
BRUCE SZELEST

Administrative Coordinator
BETH SIMPSON

Local Arrangements Chair
JOHN RYAN

AV Coordinator
SHIRLEY WONG

Dinner Groups
LAURA MASSELL
JENNIFER NUCEDER

Internet Café
J.R. BJERKLIE

LAC Web Pages
ZEKE BERNSTEIN

Meal Tickets
HOPE BAKER-CARR

Name Badges
HOPE BAKER-CARR

Session Room Checkers
ADELA LANGROCK
BETSY CARROLL

Signs
ZEKE BERNSTEIN

Program Chair
CHERRY DANIELSON

Associate Program Chair
SHIRLEY WONG

Pre-Conf Workshop Coordinator
CHAD MAY

Best Paper Coordinator
JENNIFER MAY

Conference Website Coordinator
MARIAN SHERWOOD

Evaluation Coordinator
JOEL BLOOM

Exhibitor Coordinator
SALLY FRAZEE

Guidebook Coordinator
KENNETH SMITH

Member Benefits Coordinator
ELIZABETH CLUNE-KNEUER

Poster Session Coordinator
CAROL VAN ZILE-TAMSEN

Proposal Review Coordinator
ALEXANDER YIN

Publications Coordinator
TIFFANY PARKER

www.neair.org

NEAIR
North East Association for Institutional Research
A guiding light for your institutional research community

Dear friends and colleagues,

I am pleased to introduce the *Proceedings* of the 42nd North East Association for Institutional Research (NEAIR) annual conference, held October 31, 2015 through November 3, 2015 at the Sheraton Burlington Hotel and Conference Center in Burlington, Vermont.

NEAIR makes public these *Proceedings* as means of sharing the contributions to the field by generous colleagues who have taken the time to prepare and present their good work, and who have taken the extra step to make their work publicly available for the historical record.

Just over 350 attendees gathered at the Sheraton Burlington Hotel and Conference Center over four sun filled warm and glorious days in mid-Autumn New England to learn from one another and network their way to future success. Our conference planning team, led by Cherry Danielson, Program Chair, John Ryan, Local Arrangements Chair, and Beth Simpson, NEAIR Administrative Coordinator, delivered yet another high quality NEAIR conference event to the membership.

Once again, I am pleased to report that networking and professional development satisfaction of attendees were among the highest rated areas in the conference evaluation, along with program content. These were the program team's highest priority objectives going into the 42nd conference. It is thus heartening to learn that the good efforts of the conference planning team paid off in the eyes of the membership.

We are ever indebted to Tiffany Parker, NEAIR Publications Coordinator, for making these proceedings available for the public record as a guiding light for your Institutional Research community. I hope you enjoy revisiting these *Proceedings* as much as we attendees appreciated them in person.

Sincerely,

Bruce Szelest
NEAIR co-President 2014-15

2014-2015 LEADERSHIP TEAM

Officers:

Co-Presidents	Mark Palladino, Bruce Szelest
Secretary ('14 -'17)	Jennifer Dunseath
Treasurer ('13 -'17)	Stephen Sheridan

Steering Committee Members:

President-Elect	Mark Palladino
Immediate Past President	Bruce Szelest
Program Chair	Cherry Danielson
Local Arrangements Chair	John Ryan
Member-At-Large ('12 -'15)	Cristi Carson
Technology Chair ('12 -'15)	Ingrid Skadberg
Member-At-Large ('13 -'16)	Elizabeth Clune-Kneuer
Member-At-Large ('13 -'16)	Shannon Tinney Lichtinger
Member-At-Large ('14 -'17)	K. Tracy Barnes
Member-At-Large ('14 -'17)	Chad Muntz
Administrative Coordinator (ex-officio)	Beth Simpson

STANDING COMMITTEES

Finance Committee

Chair	**Cristi Carson**
Chair-elect	Chad Muntz
Treasurer	Stephen Sheridan
Member ('12-'15)	Thomas Dahlstrom
Member ('14-'17)	Gurvinder Khaneja

Grants Committee

Chair	**Shannon Lichtinger**
Chair-Elect	TBD 2015
Immediate Past Chair	Laura Uerling
Grants Reviewer ('12-'15)	Alexa Beshara-Blauth
Grants Reviewer ('14-'17)	Seth Ovadia
Grants Reviewer ('14-'17)	Charis Ng
Grants Reviewer ('14-'17)	June Ragone

Professional Development Services (One Year Term)

Chair	**Mark Palladino**
Workshop Recruitment	Chad May
Local Arrangements	Betsy Carroll
Site Logistics & Registration	TBD

Membership

Chair	**Elizabeth Clune-Kneuer**
Chair-elect	TBD 2015
Immediate Past Chair	Maren Hess
Groups ('14-'17)	Anj Shaw
Job & News Posting ('14-'17)	Diana Danelian
Mentors ('13-'16)	Melanie Larson
Mid-Career Focus ('13-'16)	Alison Weingarten
Newcomers ('14-'17)	Jennifer May
Social Media ('14-'17)	Claire Goverts

Nominations (One Year Term)

Chair	**Bruce Szelest**
Member (Community College)	Phyllis Fitzpatrick
Member	Lee Allard
Member	Ellen Boylan
Member	Chad May
Member	Paul Prewitt-Freilino
Member	Alan Sturtz

Site Selection

Chair	**Mark Palladino**
Treasurer	Stephen Sheridan
Past LAC or PC ('13-'16)	Shannon Lichtinger
Admin Coordinator	Beth Simpson

Technology

Chair	**Ingrid Skadberg**
Chair-elect	Tracy Barnes
Website ('13-'16)	Joseph Stankovich
Website ('13-'16)	Jennifer Lewis
Website ('13-'16)	Kenneth Smith
Access/YM ('14-'17)	Melissa Thorpe

2014-2015 LEADERSHIP TEAM

Program (One Year Term)

Chair	**Cherry Danielson**
Associate Program Chair	Shirley Wong
Exhibitor ('13-'16)	Sally Frazee
Pre-Conf Workshops	Chad May
Best Paper ('13 -'16)	Jennifer May
Conf Website ('12-'15)	Marian Sherwood
Evaluation ('13-'16)	Joel Bloom
Focus Group &	Shama Akhtar
Closing Raffle	Chad May
Guidebook ('12-'15)	Kenneth Smith
Member Benefits	Elizabeth Clune-Kneuer
OpenConf ('12-'15)	Alexander Yin
Poster Session ('12-'15)	Carol Van Zile-Tamsen
Publications ('12-'15)	Tiffany Parker

Local Arrangements (One Year Term)

Chair	**John Ryan**
AV Coordinator	Shirley Wong
Dinner Group	Laura Massell
	Jennifer Nuceder
Internet Café	JR Bjerklie
LAC Web pages	Zeke Bernstein
Meal Tickets/Badges	Hope Baker-Carr
Session Room Checker	Adela Langrock
	Betsy Carroll
Signs	Zeke Bernstein

Conference Proposal Peer Reviewers

Alexander Yin, Peer Review Coordinator

Lee Allard
Andrea Bakker
Libby Barlow
Annemarie Bartlett
Robin Beads
Felice Billups
J.R. Bjerklie
Gary Boden
Heather Bouchey
Christina Butler
Jason Canales
Angela Carlson-Bancroft
Betsy Carroll
Cristi Carson
Elizabeth Clune-Kneuer
Margaret Cohen
Theodore Coladarci
Lauren Conoscenti
Dan Custer
Veena Dhankher
Stephanie Dolamore
Suhua Dong
Sarah Donnelly
Kate Doria
Jennifer Dunseath
Mark Eckstein
Krisztina Filep
Marcia Finch
Gayle Fink
Emily Foley
Sally Frazee
Nora Galambos
Brandy Gray
Rachel Groenhout
Rommel Guadalupe
Linda Heineman
Braden Hosch
Christopher Hourigan
May Hser
Michelle Kiec
Lester Ko
Elisabeth Lackner
Raldy Laguilles
Cynthia Langin
Melanie Larson
Ann Lehman
Jeannette Lim
Jack Mahoney
Laura Massell
Tom McGuinness
Edmund Melia
James (Jim) Miller
Louise Murray
Matt North
Gloria Oikelome
Katherine Ostroot
Allison Ostrowski
Suzanne Phillips
Paul Prewitt-Freilino
Heather Roscoe
Cate Rowen
Nathan Rush
Ann Marie Russell
Austin Ryland
Jessica Sharkness
Jessica Shedd
Mary Heather Smith
Kristen Sullivan
Valerie Swain
Wayne Taliaferro
Gail Tudor
Carol Van Zile-Tamsen
Christopher Vinger
Courtney Wade
Vennessa Walker
Lynn Wallace
Allison Walters
Andrew Watters
Michael Whitcomb
Shirley Wong
Alexander Yin

2014-2015 LEADERSHIP TEAM

Best Paper Committee ('14-'17)
Jennifer May, Best Paper Coordinator

Gina Calzaferri	Braden Hosch	Melanie Sullivan
Michael Dillon	Thomas McGuinness	

ACKNOWLEDGEMENT

Contained in these pages are the Proceedings of the NEAIR 42nd Annual Conference provide eleven contributed conference papers and three IR reports authored by 18 NEAIR colleagues.

Additional conference presentations are just a few clicks away – accessible within in the NEAIR website under the Annual Proceedings section. These pages are only accessible to signed in NEAIR members.

Special thanks to Bruce Szelest, Cherry Danielson, Jennifer May and Beth Simpson for their contributions, oversight, and support with all aspects of publications responsibilities during the course of this past year.

Tiffany Parker

2014-2015 NEAIR Publications Coordinator

Mt. Wachusett Community College

Table of Contents

****Indicates Best Paper Award**

*** Indicates Best First Paper Award**

STRATEGIES TO ANALYZE COURSE AND TEACHING EVALUATION DATA

Kati Li

Research Analyst, Office of Institutional Research and Assessment

Temple University

Abstract

This paper describes the steps taken by one large public university to analyze, summarize, and present key findings on its teaching evaluations. A composite score summarizing course ratings was created, and tests were conducted to evaluate whether the composite score varied by course level, college type/academic discipline, instructional method, and instructor type. Box plots showed that courses were rated highly overall. Kruskal-Wallis tests found the predictor variables of interest were significantly related to the composite score, although effect sizes were small. The paper concludes with implications for current practice and future research.

Introduction

Colleges and universities across the nation conduct course and teaching evaluations, and the data from these evaluations are important for many reasons. Course evaluations provide students an opportunity to voice their thoughts about their courses; faculty, in turn, use data from the evaluations to improve their teaching and to create better learning experiences for students. On an administrative level, course and teaching evaluations are used by departments to inform faculty promotion and hiring decisions. Typically, institutional research departments collect data on course and teaching evaluations, and they are uniquely positioned to make sense of the data on a university-wide level. Course and teaching data contain a wealth of information, and it is up to institutional researchers to sift through the data with all of its complexities and nuances and to ultimately communicate their findings in accessible, clear ways. This paper examines the process taken by an institutional research department of one large public university – Temple University – to analyze, summarize, and present key findings on its teaching evaluation data. The steps included searching the existing literature to identify relevant variables, choosing a data source, constructing variables, and running statistical analyses. Results from this study pave the way for future research on course evaluations at Temple. Additionally, as will be discussed at the end of the paper, the steps taken to conduct this particular study could be transferred and applied to other universities and colleges as they engage in their own analyses of course and evaluation data.

Literature Review

The existing literature identifies course and instructor-level variables that are associated with student ratings of teaching: course level, college type/academic discipline, instructional method, and instructor type.

Course Level

Higher-level courses, in particular graduate-level courses, are rated higher than lower-level courses (Aleamoni, 1981; Braskamp & Ory, 1994; Feldman, 1978), although these differences tend to be small. Feldman (1978) finds that the association between course level and ratings are diminished when other factors—class size, expected grade, and electivity—are added as controls. It is not clear whether the effect of course level on ratings is "direct, indirect, or both" (p. 196).

Academic Discipline/College Type

Student ratings vary by discipline: humanities and arts courses receive higher ratings than social science courses, which in turn receive higher ratings than math and science courses (Braskamp & Ory, 1994; Cashin, 1990; Centra, 1993, 2009; Feldman, 1978; Hoyt & Lee, 2002a; Marsh & Dunkin, 1992; Sixbury & Cashin, 1995). Some theories have been put forth to explain these differences: students may be less prepared for quantitative courses (Benton and Cashin, 2012; Cashin, 1990), funding and research requirements are more extensive for math and science faculty, drawing time away from teaching (Cashin, 1990), and math and science disciplines continue to change and evolve quickly, so their course content is more fluid and difficult to teach (Centra, 2009).

Instructional Method

The existing body of literature does not find a consistent pattern of differences between distance/online education and traditional forms of instruction. Item means and overall assessments of instructors are similar or identical between online and face-to-face sections (Bernard et al., 2004; Machtmes & Asher, 2000; Wang & Newlin, 2000). While some find that students express a preference for classroom education (Allen, Bourhis, Burrell, & Mabry, 2002; Ungerleider & Burns, 2003), when it comes to academic achievement, students taking courses by distance education are no different than students in traditionally-instructed courses (Ungerleider & Burns, 2003), and, in some cases, they may actually outperform the students receiving traditional instruction (Shachar & Neumann, 2003).

Instructor Type

The research on instructor type is mixed. Peters and Chow (1988) do not identify differences in teaching/course ratings by instructor type. Graduate students received lower course ratings in a study conducted by Braskamp and Ory (1994). McPherson and Jewell (2007) find that tenured professors outperform non-tenured professors; on the other hand, Feldman (1983) concludes that teaching ratings peak at 6-8 years of teaching, and then gradually decline, a pattern that coincides roughly with the tenure decisions at most institutions.

Data Sources and Methodology

Data Sources

Data was drawn from the Fall 2014 Temple University Student Course and Teaching and Evaluation (called "Student Feedback Form" at Temple). Temple University is a large public research university located in Philadelphia, Pennsylvania with 464 active programs and over 38,000 students enrolled. At Temple, course and teaching evaluations are offered both online and on paper, with the vast majority of evaluations being online. This analysis contains data from both online and paper evaluations.

There are five types of course and evaluation forms at Temple, tailored for different instructional types: (1) Basic, Single Instructor; (2) Laboratory Section; (3) Recitation or Workshop; (4) Performance or Studio-Based Courses; (5) Multiple Instructors. The Single Instructor Form and the Multiple Instructors Form contain the same questions. For the other forms, many questions overlap or are similar to the Single Instructor Form, and a few questions are form-specific. Evaluation items consist of three types: (1) questions that assess student's preparation for the course; (2) questions on the instructor's teaching; and (3) questions on the overall quality of the course. At the end of each form, students have an option to leave open-ended comments.

For this paper, data from the Single Instructor and Multiple Instructor Form (which comprised over eighty percent of all evaluation forms) were included in the analysis. For cross-listed and multiple instructor courses (for which there would be duplicate course and teaching evaluation data), the data for only one course and the first instructor was kept. In course sections with 4 or less responses, the presence of one or two extreme values could easily bias the average ratings, so those sections were eliminated from the analyses.

Outcome Variable. Temple's course and teaching evaluation forms cover a large number of items, so to streamline the analysis, a composite score variable was created that combined four items of the course and teaching evaluation: (1) The instructor provided useful feedback about exams, projects, and assignments; (2) So far, the instructor has applied grading policies fairly; (3) The instructor taught this course well; and (4) I learned a great deal in this course. The composite score ranged from 1 to 5, with higher scores indicating more favorable assessments of the course: 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree.

The composite score variable was weighted to account for differences in course sizes and was calculated as follows:

$$\text{Composite Score} = \frac{(q1n * q1m) + (q2n * q2m) + (q3n * q3m) + (q4n * q4m)}{(q1n+q2n+q3n+q4n)}$$

n = number of responses, m = mean score of course section

Predictor Variables. Course level was categorized as follows: Preparatory (referring to 700-level courses that students take in preparation for more advanced college-level courses), General Education (Gen Ed), Lower Division, Upper Division, and Graduate/Professional.

Temple University is comprised of several schools and colleges that represent different academic disciplines. For this paper, the 18 schools/colleges were re-coded into the following categories: (1) Humanities; (2) Social Sciences; (3) Professional; (4) Science/Math; (5) Other.

Three kinds of instructional methods were examined: Classroom (courses taught face-to-face); Online/Video/Virtual/Hybrid; and Other/Unknown. Online, video, virtual and hybrid courses were combined into one category to ensure an adequate sample size.

The following instructor types were examined: Graduate Student; Adjunct; Tenure-Track, Tenured, Non-Tenure-Track, and Other/Unknown. Non-Tenure-Track faculty are faculty who work full-time and are not adjuncts, tenured, tenure-track, or graduate students.

Methodology

Descriptive statistics - minimums, maximums, 25th percentile scores, 50th percentile scores (medians), 75th percentile scores, and means – were calculated and presented as tables and box plots. Box plots were included to provide a visual depiction of course rating distributions. The bottom of each box marks the 25th percentile score, the top of the box is the 75th percentile, and the line in the middle represents the 50th percentile score. The bottom horizontal stroke of the box plot demarcates the minimum value, and the top horizontal stroke of the box plot is the maximum value. Means are shown as dots inside of the box plots.

Box plots provided insight into the general distribution of the data, but to identify statistically significant differences, more rigorous tests were needed. The Kruskal-Wallis test determines if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable and was an appropriate test for this study.[1] All of the predictor variables—course level, college, instructional method, and instructor type— had two or more groups, and the outcome variable, the composite score, was an ordinal (ranked) variable.

The Kruskal-Wallis tests determines whether there are differences between the groups of a predictor variable, but not which groups are different. To assess differences between the categories of predictor variables (for example, do higher level courses outperform lower level courses or do humanities courses rate higher than math/science courses), post-hoc Mann-Whitney U tests using Bonferroni correction were run.

Just because something is statistically different does not necessarily mean that it has practical or theoretical significance. For example, lower division courses could have ratings of 4.0 and higher division courses could have ratings of 4.2, but since the ratings are scaled from 1 to 5, a 4.0 means essentially the same thing as a 4.2: both course types have performed exceedingly well. Thus, to identify whether statistically significant differences had any practical or theoretical significance, the effect size (also known as 'strength of association') was calculated and evaluated using Cohen's (1988) criteria of .1 = small effect, .3 = medium effect,

[1] In supplemental analyses, an ANOVA test was run and produced similar results as those obtained with the Kruskal-Wallis test. The results of the Kruskal-Wallis are presented in this paper for two reasons: (1) to account for the ordinal (ranked) nature of the course evaluation ratings; (2) to err on the side of caution (the Kruskal-Wallis is a non-parametric test that makes less assumptions than the ANOVA and produces more conservative results).

.5 = large effect. The formula to calculate effect size was as follows: r = z / square root of N, where N = total number of cases.

Results: Box Plots and Tables of Descriptive Statistics

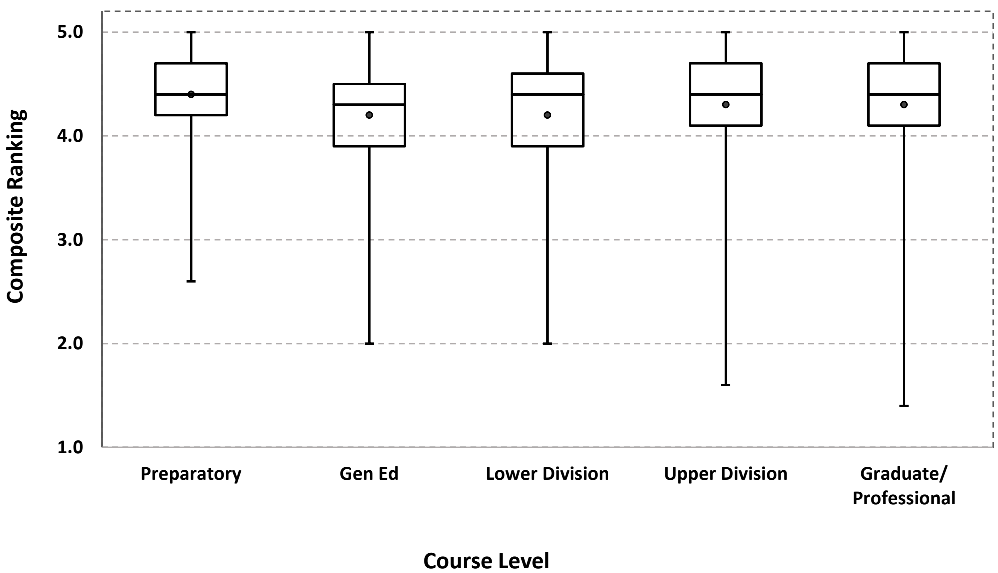


Figure 1. Box plot depicting composite ranking by course level.

Table 1

Descriptive Statistics of Composite Score by Course Level.

	Preparatory	Gen Ed	Lower Division	Upper Division	Graduate/ Professional
Minimum	2.6	2.0	2.0	1.6	1.4
25th Percentile	4.2	3.9	3.9	4.1	4.1
Median	4.4	4.3	4.4	4.4	4.4
75th Percentile	4.7	4.5	4.6	4.7	4.7
Maximum	5.0	5.0	5.0	5.0	5.0
Mean	4.4	4.2	4.2	4.3	4.3

Figure 1 and Table 1 show that there is a tendency toward agreement or strong agreement on the composite teaching evaluation score by course level. The top 75% of ratings for Preparatory, Upper Division, and Graduate/Professional courses were 4.0 or above, and 75% of ratings for Gen Ed courses and Lower Division courses were 3.9 or above. Mean ratings were 4.4 for Preparatory, Lower Division, Upper Division, and Graduate/Professional courses and 4.3 for Gen Ed courses. For all course levels, the maximum course rating was 5.0. The lowest minimum rating by course level was 1.4 (Graduate/Professional) and the highest minimum rating by course level was 2.6 (Preparatory).

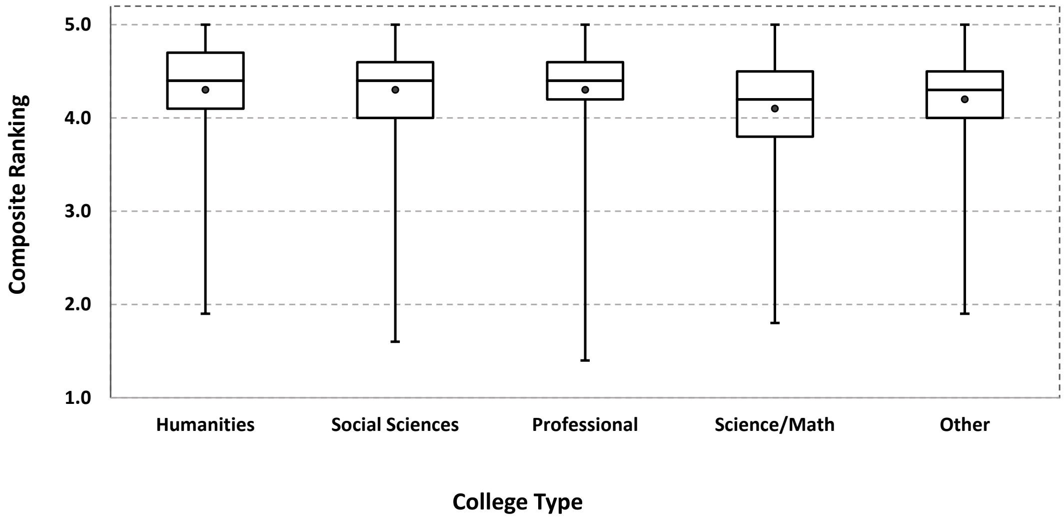


Figure 2. Box plot showing composite ranking by college type.

Table 2

Descriptive Statistics of Composite Score by College Type.

	Humanities	Social Sciences	Professional	Science/Math	Other
Minimum	1.9	1.6	1.4	1.8	1.9
Percentile 25	4.1	4.0	4.2	3.8	4.0
Median	4.4	4.4	4.4	4.2	4.3
Percentile 75	4.7	4.6	4.6	4.5	4.5
Maximum	5.0	5.0	5.0	5.0	5.0
Mean	4.3	4.3	4.3	4.1	4.2

As shown in Figure 2 and Table 2, across college types, there was a tendency toward agreement or strong agreement on the composite teaching evaluation score. The top 75% of ratings for Humanities, Social Sciences, Professional, and Other colleges was 4.0 or above. Median ratings were 4.4 for Humanities, Social Sciences, and Professional colleges, 4.2 for

Science/Math colleges, and 4.3 for Other colleges. For all college types, the maximum composite score was 5.0. The lowest minimum composite score was 1.4 (Professional colleges).

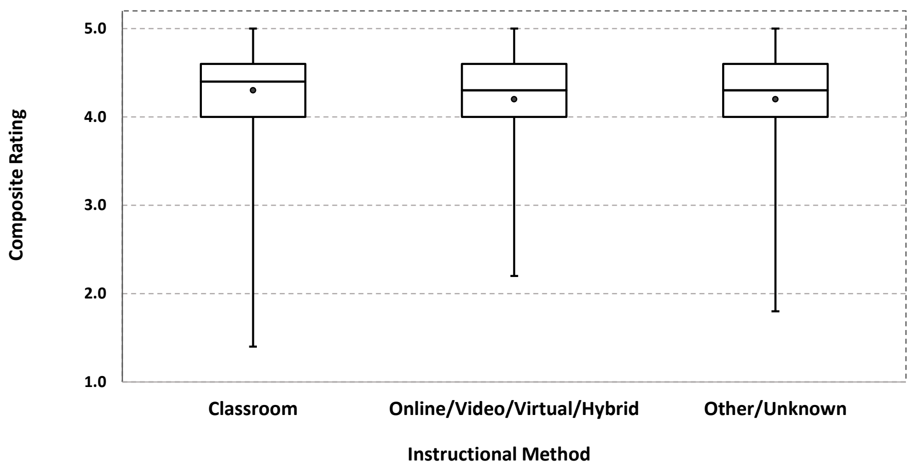


Figure 3. Box plot of composite ranking by instructional method.

Table 3

Descriptive Statistics of Composite Score by Instructional Method.

	Classroom	Online/Video/ Virtual/Hybrid	Other/Unknown
Minimum	1.4	2.2	1.8
25th Percentile	4.0	4.0	4.0
Median	4.4	4.3	4.3
75th Percentile	4.6	4.6	4.6
Maximum	5.0	5.0	5.0
Mean	4.3	4.2	4.2

According to Figure 3 and Table 3, across instructional methods, there was a tendency toward agreement or strong agreement on the composite score. The top 75% of ratings for

Classroom, Online/Video/Virtual/Hybrid, and Other/Unknown instructional methods were 4.0 or above. Median ratings were 4.4 for Classroom methods, and 4.3 for Online/Video/Virtual/Hybrid methods and Other/Unknown methods. For all instructional methods, the maximum course rating was 5.0. The lowest course rating was 1.4 (Classroom).

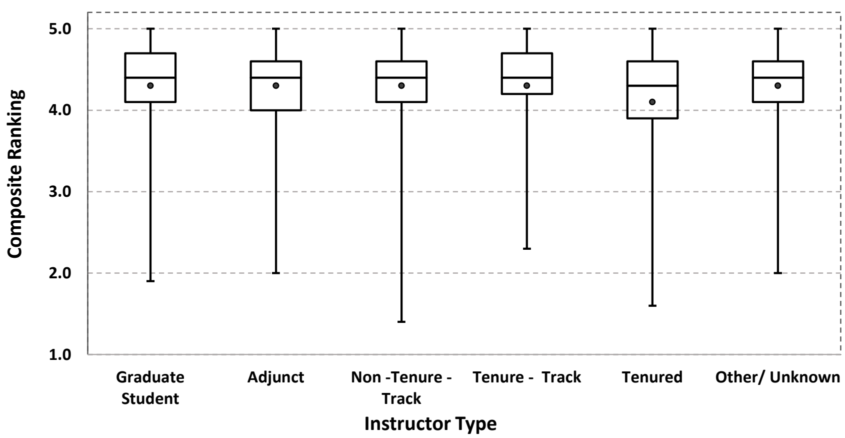


Figure 4. Box plot of composite ranking by instructor type.

Table 4

Descriptive Statistics of Composite Score by Instructor Type.

	Graduate Student	Adjunct	Non -Tenure- Track	Tenure - Track	Tenured	Other / Unknown
Minimum	1.9	2.0	1.4	2.3	1.6	2.0
25th Percentile	4.1	4.0	4.1	4.2	3.9	4.1
Median	4.4	4.4	4.4	4.4	4.3	4.4
75th Percentile	4.7	4.6	4.6	4.7	4.6	4.6
Maximum	5.0	5.0	5.0	5.0	5.0	5.0
Mean	4.3	4.3	4.3	4.3	4.1	4.3

As shown in Figure 4 and Table 4, across instructor types, there was a tendency toward agreement or strong agreement on the composite teaching evaluation score. The top 75% of ratings for Graduate Students, Adjuncts, Non-Tenure-Track, Tenure-Track, and Other/Unknown instructors were 4.0 or above, and 75% of ratings for Tenured instructors was 3.9 or above. Median ratings averaged 4.3 for Tenured instructors and 4.4 for the other instructor types. For all types, the maximum course rating was 5.0. The lowest course rating was 1.4 (Non-Tenure-Track instructors).

Results: Kruskal-Wallis, Mann-Whitney U, and Effect Sizes

Course Level

A Kruskal-Wallis test was conducted and showed that the composite score varied significantly by course level, χ^2 (4, n = 4366) = 52.02, p = .000. Follow up Mann-Whitney U tests were conducted to evaluate pairwise differences among the five groups, controlling for Type I error across tests by using the Bonferroni correction. The results of these tests revealed significant differences in the composite score of Gen Ed (Md = 4.3, n = 852) and the other course levels: Graduate/Professional (Md = 4.4, n = 735), U = 258877, z = -5.96, p = .000, r = .15; Preparatory (Md = 4.4, n = 104), U = 34206, z = -3.80, p = .000, r = .12; Lower Division (Md = 4.4 , n = 796), U = 311867, z = -2.82, p = .005, r = .07; and Upper Division (Md = 4.4, n = 1879), U = 681216, z = -6.25, p = .000, r = .12. In summary, the Gen Ed courses were rated significantly lower than the other course types, but these differences were of relatively little practical or meaningful significance.

College Type

A Kruskal-Wallis test revealed significant differences among the five college types (Humanities, Social Sciences, Professional, Science/Math, Other) on median change in the composite score, χ^2 (4, n = 4366) = 114.18, p = .000. Mann-Whitney U tests using Bonferroni correction identified significant differences between Science/Math (*Md* = 4.2, n = 584) and the following college types: Humanities (*Md* = 4.4, n = 2155), U = 455960, z = -10.22, p = .000, r = .20; Social Sciences (*Md* = 4.4 , n = 1297), U = 292486, z = -7.91, p = .000, r = .18; and Professional (*Md* = 4.4 , n = 169), U = 34004, z = -6.16, p = .000, r = .22. The Other college type (*Md* = 4.3, n = 101) rated significantly lower than Humanities (*Md* = 4.4, n = 2155), U = 90039, z = -2.94, p = .003, r = .06 and Professional (*Md* = 4.4, n = 169), U = 6767, z = -2.85, p = .004, r = .17. Although the Science/Math and Other courses had lower ratings, the effect sizes were small, revealing that the differences were of low theoretical or practical significance.

Instructional Method

A Kruskal-Wallis test that was conducted to evaluate differences in the three instructional methods (Classroom, Online/Video/Virtual/Hybrid, Other/Unknown) was marginally significant χ^2 (2, n = 4366) = 6.64, p = .036. Follow-up Mann-Whitney U tests using Bonferroni correction revealed marginally significant differences in the composite score of Classroom methods (*Md* = 4.4, n = 3925) and Online/Video/Virtual/Hybrid methods (*Md* = 4.3, n = 175), U = 310110, z = -2.18, p = .030. The effect size was very small, r = .03, meaning that in practice, the Classroom methods and Online/Video/Virtual/Hybrid courses were not altogether that different. Both instructional methods performed very well, scoring between an “agree” and “strongly agree” on the composite score.

Instructor Type

A Kruskal-Wallis test revealed a statistically significant difference in the composite score across the six instructor types, χ^2 (5, n = 4366) = 36.43, p = .005. Mann-Whitney U tests using Bonferroni correction identified significant differences between the Tenured faculty (Md = 4.3, n = 782) and all other instructor types: Adjunct (Md = 4.4 , n = 1254), U = 433299, z = -4.42, p = .000, r = 10; Graduate Student (Md = 4.4, n = 258), U = 85790, z = -3.61, p = .000, r = .11; Non-Tenure-Track (Md = 4.4 , n = 1369), U = 465660, z = -5.03, p = .000, r = .11; Tenure Track (Md = 4.4 , n = 225), U = 70673, z = -4.50, p = .000, r = .14 ; and Other (Md = 4.4, n = 478), U = 164820, z = -3.52, p = .000, r = .10. In essence, tenured instructors were rated lower than the other instructor types, but the differences were relatively small.

Summary of Results and Relationship to Existing Literature

The results of the Kruskal-Wallis, Mann-Whitney U, and Effect Size tests are summarized below in Table 5.

Table 5

Kruskal-Wallis, Mann-Whitney U, and Effect Sizes of Temple Fall 2014 Course Rating Data.

Predictor Variable	Kruskal-Wallis	Mann-Whitney U	Effect Size
Course Level	Significant (p = .000)	Gen Ed rated lower than other course levels	Small (r = .12)
College Type	Significant (p = .000)	Science/Math lower than Humanities, Social Sciences, and Professional Other rated lower than Humanities and Professional	Small (r = .17)
Instructional Method	Significant (p = .036)	Classroom rated slightly higher than Online/Video/Virtual/Hybrid	Very Small (r = .03)
Instructor Type	Significant (p = .005)	Tenured rated lower than the other instructor types	Small (r = .10)

Results from this study can be situated within the broader literature on course evaluations. In general, this study finds that higher-level courses are rated higher than lower-level courses, a pattern that is consistent with other studies, although one notable difference appeared: preparatory courses did very well, at levels comparable to the upper-level courses. Also consistent with the past literature, science/math courses tended to be rated lower than the other academic disciplines. Similar to how other studies find no consistent pattern of course rating differences by instructional method, this study found that classroom methods did not achieve practical or theoretical significance. So far, the research on instructor type has produced mixed results; in this study, it appears that the tenured professors received slightly lower ratings.

Conclusions - Plans for Future Research

Three research topics emerge from the results of this study: First, preparatory courses performed exceptionally well, exceeding expectations based on the existing literature. It would be worthwhile to explore and identify the factors that explain the success of Preparatory courses and to perhaps replicate these strategies with other course types. A mixed methods approach might be best: multivariate statistical analyses that test for mediating and moderating variables could be combined with interviews of preparatory course faculty and in-class observations of faculty methods and strategies.

Second, more research could be done to investigate why courses in the math/sciences have lower ratings. The existing literature on student ratings and academic discipline may provide a helpful starting point: It could be that instructors in fields requiring more quantitative reasoning skills are rated lower because today's students have less preparation/training in those skills (Benton & Cashin, 2012; Cashin, 1990). A second possibility is that math and science teachers may spend more of their time seeking funds and doing research time than teaching, relative to their humanities/social science counterparts (Cashin, 1990). It may also be worth exploring whether natural science courses may be more difficult to teach because knowledge is growing more rapidly in those areas and teachers feel pressured to cover increasing amount of material; as a result, students find learning the material more challenging (Centra, 2009).

Third, research could investigate the lower ratings of Tenured instructors. Since the existing literature on instructor rank is mixed, it is worth investigating if differences by instructor rank found in this study are driven by interrelated variables. For example, courses with larger class sizes tend to receive poorer ratings (Aleamoni & Hexner, 1980; Centra, 2009; Hoyt & Lee, 2002b); so if Tenured instructors are more likely to teach larger classes than other instructor types, the significant differences for Tenured professors may be driven largely by course size. Adding class size as well as other control variables in multivariate analyses and testing for interaction effects may lend insight into disentangling the relationship between instructor type and course ratings.

Conclusions - Implications for Current Practice

The steps taken for this study provide a model of analyzing and presenting course evaluation data that can be applied to institutional research departments at other universities. When it comes to constructing variables and preparing the data for analysis, it may help to create a composite core that combines key items of the course evaluation. This strategy enables institutional researchers to streamline their analyses and allows the audience to quickly make sense of the results because there is just one outcome variable to focus on. Second, since course sections vary in size, it may be worthwhile to (1) delete from the analyses any courses with very few responses (in the case of this paper, courses with 4 or less responses were removed from analyses) and (2) weight the composite score. Using these approaches, larger courses and courses with more responses account for a greater share of the overall results.

The analysis and presentation of course evaluation data should involve two key steps. The first step is to gain a general understanding of the distribution of the data. Besides using tables, it is worth considering using box plots to provide a visual representation of the data and to estimate differences between categories. Box plots communicate a clear and compelling message about the data that is easier to decipher than a table of values. The second step is to assess statistical significance and effect size and to consider the results of the tests jointly in the final evaluation of the data. The results of this study showed that, although General Education courses, Math/Science courses, and courses taught by Tenured instructors had lower ratings, these differences had little practical or meaningful difference. Conclusions and recommendations that follow from this analysis should take into account that the groups with lower ratings had high ratings overall.

References

Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 110-145). Beverly Hills, CA: Sage.

Aleamoni, L. M., & Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science, 9*, 67-84.

Allen, M., Bourhis, J., Burrell N., & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. *American Journal of Distance Education, 16*(2), 83-97.

Benton, S. L., & Cashin, W. E. (2012). *IDEA Paper No. 50: Student ratings of teaching: A summary of research and literature.* Manhattan, KS: The IDEA Center.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., and Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Braskamp, L. A, & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.

Cashin, W. E. (1990). Students do rate different academic fields differently. In M. Theall and J. Franklin (eds.), Student Ratings of Instruction: Issues for Improving Practice, *New Directions for Teaching and Learning*. 43. San Francisco: Jossey-Bass.

Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness.* San Francisco: Jossey-Bass.

Centra, J. A. (2009). *Differences in responses to the Student Instructional Report: Is it bias?* Princeton, NJ: Educational Testing Service.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teacher: What we know and what we don't. *Research in Higher Education*, *9*, 199-242.

Feldman, K. A. (1983). The seniority and instructional experience of college teachers as related to the evaluations they receive from their students. *Research in Higher Education, 5*(3), 243-288.

Hoyt, D. P., & Lee, E. (2002a). *Technical Report No. 13: Disciplinary differences in student ratings.* Manhattan, KS: The IDEA Center.

Hoyt, D. P., & Lee, E. (2002b). *IDEA Technical Report No. 12: Basic data for the revised IDEA system.* Manhattan, KS: The IDEA Center.

Machtmes, K., & Asher, J. W. (2000). A meta-analysis of the effectiveness of telecourses in distance education. *American Journal of Distance Education*, *14*(1), 27-46.

Marsh H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 8). New York: Agathon Press.

McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly, 88*(3), 868-881.

Peters, M. K., & Chow, J. C. (1988). Sources of variation in students' evaluations of instruction in a graduate social work program. *Journal of Social Work Education, 24*(1), 35-42.

Shachar, M., & Neumann, Y. (2003). Differences between traditional and distance education academic performances: A meta-analytic approach. *International Review of Research in Open and Distance Education, 4*(2), 1-20.

Sixbury, G. R., & Cashin, W. E. (1995). *IDEA technical report No. 10: Comparative data by academic field.* Manhattan: Kansas State University, Center for Faculty Evaluation and Development.

Ungerleider, C., & Burns, T. (2003). *A systematic review of the effectiveness and efficiency of Networked ICT in Education*. A State of the Field Report to the Council of Ministers of Education, Canada, and Industry Canada. Ottawa, ON: Industry Canada.

Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, *92*, 137-143.

USING A MIXED METHODS APPROACH TO

ASSESS A LEADERSHIP MENTORING PROGRAM

Betty J. Harper

Interim Executive Director, Office of Planning and Institutional Assessment

The Pennsylvania State University

Abstract

Penn State's Administrative Fellows Program has provided a prototype for faculty and staff mentoring around the country. In 2014-15, the University conducted a mixed-methods evaluation of that program. Data was collected via survey, focus groups, and interviews. The findings highlights its outcomes and the benefits derived to participants and the University. The findings provide an overview of the typical Fellow's experience, identifies strengths and weaknesses of the program, and identifies potential ways in which this and similar programs might be improved. Further, this project highlights the benefits and costs of mixed-methods approaches to program evaluation for institutional researchers.

Introduction

In our quest to support the growth and development of faculty, students, and staff, colleges and universities often implement interventions that rely on significant monetary and human resources and benefit relatively small numbers. The long-term impact of such programs can be enormous, but the evaluation of such resource-intensive programs is often neglected because of the difficulties involved in assessing and documenting those impacts. Penn State's Administrative Fellows Program (AFP) is one such program. The AFP provides faculty and staff with a one-of-a-kind year-long opportunity to be mentored by Penn State's leading executives and to observe decision making at the highest level. Women and minorities are particularly recruited and encouraged to apply.

The AFP is unique in that it has had not one, but two, comprehensive mixed-methods evaluations since its inception in 1986. This paper presents the findings from the most recent evaluation of the AFP, In addition, it highlights methodological efficiencies for institutional researchers (IR) and the benefits derived from the use of interviews and focus groups in addition to a standard program evaluation survey.

Background

What is Mentoring and Why Does it Matter?

There are numerous conceptualizations of mentoring, but this study uses Ragins and Scandura's (1999, p. 496), which focuses on mentors as "influential individuals with advanced experience and knowledge who are committed to providing upward mobility and support to their protégés' careers." Numerous studies have affirmed the importance of mentoring, particularly in the career development of women and minorities (see for example, Claire, Hukai, & McCarty, 2005; Cox & Salsberry, 2012; and Touchton, Musil & Campbell, 2008). While it is worthwhile to note that the mentoring literature has been criticized for focusing exclusively on the benefits of mentoring and ignoring drawbacks (Carr & Heiden, 2011), the well-established benefits appear to outweigh potential obstacles, such as dysfunctional mentor/protégé relationships. In their review of the literature, Blake-Beard, Murrel and Thomas (2006) noted that benefits related to mentoring include higher salaries, career advancement, career satisfaction, and institutional loyalty. In particular, mentoring relationships can play a critical role in facilitating professional promotion for individuals who face historical and cultural barriers to advancement (Baltodano, Carlson, Jackson, & Mitchel, 2012).

Despite the importance of mentors in professional development, influential mentors can be hard to find and not all have equal access to high-level mentors. While women are more likely than men to say that they've had a mentor (Ibarra, Carter, & Silva, 2010), women and minorities may have less access to influential mentors than their White, male colleagues (Dreher, & Cox, 1996; Sandberg, 2013). For this reason among others, many organizations have implemented formal mentoring programs focused on developing a diverse leadership pool. Seventy-one percent of Fortune 500 companies offer mentoring programs for their employees (Chronus, 2012) and colleges and universities are increasingly offering formal mentoring programs designed to develop future administrators.

Administrative Mentoring Programs

Formal mentoring programs for both students and tenure-track faculty are common in higher education, but mentoring programs geared toward administrative leadership are less so. The most well-known administrative mentoring programs are found in academic hospital settings (e.g., Johns Hopkins and the Mayo Clinic's Administrative Fellowship Programs). These programs focus on early career development by introducing entering professionals to administrative roles. Formal mentoring programs can be found at other colleges and universities (for example, Iowa State, Ohio State, and Purdue) and in higher education organizations such as the Committee on Institutional Cooperation and the Southeastern Conference.

Program Structure

In the AFP, three Mentors are recruited annually from among the University's provost and the vice presidents. Fellowship applicants must hold a standing, full-time faculty or staff appointment and may be located at any Penn State location, but must be willing to spend the Fellowship year at the Mentor's campus (most Mentors are located at the University Park campus). A steering committee reviews applications, conducts preliminary interviews, and provides recommendations to the Mentors. Mentors interview a short-list of prospective Fellows and make the final decision.

In order to help minimize the disruption inherent in removing a faculty or staff member from their home unit for a year, each sending unit is provided funds to backfill the Fellow's position. The expectation is that Fellows will separate completely from their home units for the year of the Fellowship and then return to their existing positions at the end of the year with a better understanding of the complexity of higher education, an increased ability to contribute to the work of their home unit, and improved prospects for advancement.

Need for the Program

Developing leaders from within is an important component of succession planning. Internal hires hit the ground running, are less expensive, and more likely to remain than external recruits (Bidwell, 2011). Internal leadership development also increases employee engagement and retention (Lamoureaux, 2013). Outstanding leadership is not homogenous leadership. It is diverse in perspective, background, and thought (Morrison, 1992). Significant attention has been given to the growing diversity of the U.S. population and its significance in terms of student and faculty diversity in higher education. Women make up a growing majority of undergraduate students (Peter & Horn, 2005) and soon, minority students will make up nearly half of all public high-school graduates (Prescott & Bransberger, 2012). Despite the changing face of the student body, diversity lags among university faculty and administrators. In 2011, minority students made up 39% of national higher education enrollments, but only 20% of all full-time instructional faculty, 15% of senior faculty, and 20% of full- and part-time administrators (National Center for Education Statistics (NCES), 2013). Likewise, women made up 57% of the national student body, but only 29% of senior faculty and 53% of administrators. Administrators, using NCES categories, include all managerial-level staff.

In 1985, the AFP was created in response to a lack of upward mobility and underrepresentation of women and minorities among Penn State's leadership. Today, despite a strong focus on workforce diversity, cultural inclusiveness, and employment equity across higher education, its leadership remains largely homogenous in terms of race, ethnicity, and gender. The university's senior leadership, as represented together by the President's Council (president, provost, and vice presidents) and Academic Leadership Council (chancellors, deans, and vice provosts), is 34% female and 10% minority. Looking at the institution's leadership over the decade since the last evaluation of the AFP, there has been an increase in the diversity of executives, administrators, and academic administrators (Figure 1), but it is still not reflective of the diversity of Pennsylvania, which is 51% female and 17% non-White (Pennsylvania State Data Center, 2015), nor of the student body. In contrast, nearly two-thirds (62%) of Penn State's non-administrative staff positions have been and continue to be held by women. Eight percent of these positions are held by minorities.

Description of the Study

Formal mentoring programs are increasingly common in higher education, but evaluation of such programs is often cursory or lacking entirely. One of the reasons for this may be the positionality of such programs under provosts or vice presidents whose units are largely focused on academic programs (including their assessment), but not as attuned to professional development as would be, for example, a human resources unit. When such programs are administered at the highest levels of the university, institutional research offices may be called upon to evaluate them. Institutional researchers, particularly those at larger institutions, are heavily reliant on quantitative methods (Ducharme, 2014). One strength of such methods is to provide evaluators with a picture of "what" is happening in a given a program. However these methods often leave the "why" question unanswered (Howard & Borland, 2007). Using qualitative approaches in combination with quantitative approaches allows IR professionals to conduct a program evaluation that answer both the "what" and the "why" questions, in order to facilitate program improvement. Several key research questions guided this program evaluation:

1. What is the typical Fellow's experience?
2. What are the strongest aspects of the program?
3. What aspects could be improved?
4. Is the program meeting its goals?

This project applied a non-experimental, ex post facto mixed-methods case study approach, which utilized interview, focus group, and survey-based data collection (Krathwohl, 1998). Mixed methods research brings quantitative and qualitative approaches to bear on a research question. Mixed methods approaches have been gaining recognition since the 1980s, but are largely underutilized in institutional research. In his 2007 volume, *Using Mixed Methods in Institutional Research,* Richard Howe emphasized the complementarity of quantitative and qualitative approaches in IR. Nearly a decade later, mixed methods remain an underutilized IR tool. In 2014, only two Association for Institutional Research (AIR) National Forum paper presentations and one *New Directions for Institutional Research* article applied a mixed methods approach (AIR, 2014; Wiley, 2015).

In this project, a survey instrument provided an efficient way to reach out to all of the prior Fellows in the population of interest and to collect information about their experiences, perceptions, and outcomes. Interviews and focus groups allowed for more open-ended inquiries, requests for clarification, and follow-up questions that complemented the quantitative information with rich detail and explanation. Integrated, the findings from both methods provide a more holistic picture of the experiences of the Fellows, the strengths and weaknesses of the programs, and the program outcomes. In order to minimize the workload created by such an ambitious project, a staff member from human resources and a higher education doctoral student collaborated on the project. This approach brought together two offices that do not normally work together to pursue a topic of common interest. It not only lightened the IR workload, but also brought multiple disciplinary perspectives to bear on the project.

Researchers conducted individual interviews with past Mentors, Fellows, and the AFP program administrator. Recent members of the steering committee were given the option of participating in an individual interview or in a focus group discussion. Interviews and focus groups were conducted by three researchers following a shared protocol and the format was semi-structured, allowing new issues to emerge as a result of the information shared by the interviewee. Additional information about the Fellows' experience was collected through a survey that was distributed to all of the Fellows in the study population. Additional data on the career progress of the past Fellows was collected via web searches.

The study population included Fellows and Mentors from the past decade. Since 2004, 31 Fellows and 13 Mentors have participated in the AFP. The Fellows population is 67% White female, 15% minority[1] female, and 19% minority male. A selection of individuals involved in running the program and selecting and recruiting participants, termed Committee Members, were also included. Subjects were invited to participate in the study by Penn State's Vice Provost for Academic Affairs. Both Mentors and Fellows are strongly invested in the AFP and the participation rate for the study was high (Table 1). Roughly two-thirds of the Fellows and all but one of the Mentors invited to interview did so; all of the Fellows invited to participate in the survey did so.

Table 1. *Participation Rates*

Mode	**Invited**	**Participated**	**Rate of Participation**
Fellows interview	19	12	63%
Mentors interviews	8	7	88%
Fellows survey	19	19	100%
Committee interview or focus group	9	8	89%

Descriptive statistics were used to aggregate the quantitative data. The interview and focus group were analyzed using thematic analysis (Guest, MacQueen, & Namey, 2012). Like grounded theory, this approach focuses on themes that emerge from the data and is inherently inductive.

[1] Minority Fellows were Black, Hispanic, and American Indian. Two Fellows were of undeclared race.

Unlike grounded theory, the goal of this approach is on providing data that can be used to inform decision making, rather than on developing or building theory.

Each transcript was read multiple times and coded in an iterative process during which codes were refined (e.g., little-used codes collapsed and new codes identified). The analyst identified themes and triangulated findings using theory and the multiple data sources. The validity of the findings was supported using peer review and member checking with study participants (Lincoln & Guba, 1985).

Findings

The Program Experience

"It was absolutely a wonderful thing for them to invest in us in that way." Sandra expressed the overwhelmingly positive perception that past Fellows have of the program. If given the chance, most Fellows would do it again and they would recommend it to their colleagues. Fellows greatly appreciated the University's commitment to their development and saw it as an investment in the University's future. Mentors were more tempered in their enthusiasm, but were still positive about its organization and its role in leadership development. Given the positive nature of participants' experiences and observations, the primary theme that emerged from both Mentor and Fellow interviews was how to make a good program better.

Working with a Mentor.

Before undertaking any endeavor, it is as important to know where you want to go as it is to know how you will get there. Participants in the AFP chose the program because they believed it would help achieve their goals. Some Fellows entered the program with specific goals in mind (e.g., preparing to be a strong candidate for a particular job), but others did not. Regardless of where Fellows begin, "clarifying and articulating learning goals is indispensable to the success of a mentoring relationship" (Zachary & Fischler, 2011). Each Mentor approached the goal-setting experience in a unique way. Shirley recalled:

> *[My Mentor] said, 'Let me create for you the kind of environment that you need to achieve your personal goals, but we know you are going to be a significant contributor to our organization.' And that was, that was just an incredibly mind-blowing thing for him to say.*

The one constant was that the Mentors saw it as the Fellows' responsibility to make productive use of the year. To maximize their success, Fellows should be independent, motivated learners. Michelle noted, *"Your Mentor isn't going to do this for you. You have to do it yourself."*

The first few weeks of the Fellowship offer an important opportunity for the Mentor and Fellow to work together to establish preliminary goals for the year. Fellows' experiences suggest that this is not happening in a consistent and structured manner. Most Mentors and Fellows did not engage in formal goal-setting or planning activities; however they did typically begin with a frank conversation about the Fellows' expectations and the Mentors' suggestions for achieving them. Formal meetings between the Mentor and Fellow varied from weekly to monthly.

All Mentors included their Fellows in their senior staff meetings and encouraged them to meet individually with all of the units' senior staff. Fellows were provided with access to the Mentors' calendars and were permitted to attend, at their choosing, all but the most confidential discussions. Mentors felt that the most important things they could do for a Fellow was to provide access and to be candid and honest. In return, Mentors wanted their Fellows to be enthusiastic, engaged, and trustworthy. Some best practices included:

- Providing Fellows with context and expectations prior to meetings and/or debriefing with them afterwards (time permitting – no one did this every time)

- Having explicit, periodic conversations about the Fellows' progress toward their goals
- Including Fellows in various service activities outside of the University, such as attending national meetings where the Mentor was presenting

Travel time emerged as an important, informal meeting time for Mentors and Fellows. Whether it was time spent in cars and airports or simply walking across campus to attend a meeting, these unscheduled moments provided unique opportunities for Fellows to speak candidly with their Mentors. Talking about the importance of this, Shirley recalled, "*[My Mentor] and I traveled a lot together . . . He was always asking me questions and it was those questions that helped me to frame and to further fine tune what my goals were.*"

Meetings, activities, and events.

A core educational component of the AFP experience is attending meetings –committees, task forces, and leadership. In addition, Fellows are encouraged to schedule one-on-one meetings with a wide variety of University leaders to learn about their units and their roles, and to attend University-wide events and leadership development activities. Survey findings revealed that some activities are engaged in by all Fellows, while others are less universal. For example, 100% of survey respondents indicated that they had attended meetings of the President's Council, Board of Trustees, and Faculty Senate. Interestingly, although a number of Fellows interviewed for this project expressed a desire for a more formalized "curriculum" including practical workshops, Fellows did not attend the formal programs that were available to them, but not required, at a high rate. For example, only 16% reported attending the Penn State Emerging Leaders Program and none indicated that they took advantage of the Excellence in Management series (a list of recommended activities is available at in the *Guidelines for Administrative*

Fellows and Mentors at http://www.psu.edu/vpaa/pdfs/admin%20fellow/guidelinesfellows.pdf). There are a number of potential reasons for this, including timing, travel requirements, lack of communication about such opportunities, and perceptions about the utility of such programs, but this study did not address those questions. Moving forward, this could be an important area for additional research.

Engagement with other Fellows.

A number of Fellows felt that an important aspect of their Fellowship year was their engagement with other Fellows. Although the current typical cohort of three is small, the opportunity to learn from other participants was significant for many, and several noted that sharing office space facilitated that exchange. Brenda recalled, *"Sharing on office with the other Fellows] was a wonderful opportunity because. . . . I got the opportunity to see what they went through, but also to participate in the meetings and functions that they were involved in."* Fellows that were not at University Park full-time or who did not share office space had less cohort interaction, and expressed disappointment at missing this valuable learning opportunity.

Projects.

Many Fellows worked on one or more significant projects during their Fellowship year and perceptions of the utility of these projects varied. While the wide range of meetings attended by Fellows provides breadth of experience, projects are a mechanism to provide depth in a specific area. As in discussions of the college curriculum, the optimal balance between breadth and depth is debatable. For some Fellows, projects provided an important way to feel like active and contributing members of the Mentor's staff. Committee Member Ruth noted that projects gave them something to *"sink their teeth into and feel that the things they are learning, they could apply"*. This tangible task helped many Fellows to combat the lack of direction they felt.

Some Mentors also saw projects as a method to give Fellows an opportunity to use their skills and contribute to the unit. In discussing how he approached the possibility of a project with Fellows, one Mentor described the conversation in the following way:

> *I say, 'Look, you shouldn't feel guilty about this [being in the Fellowship]. If you want to, after you get to know the organization a little bit, if you want to sink your teeth into a couple of different places so you have some sort of project you are working on. . . that's fine.' But I think there is a little bit of guilt sometimes, about 'Gee, I don't feel like I am contributing now to Penn State like I was in my old role.'*

While some Mentors and Fellows saw projects as critical components of the Fellowship experience, others saw them as a distraction. When asked by her Mentor if she wanted to take on a project, Nancy responded, *"You know, for heaven's sake, I have done projects for all my life. No, I want to take this year just to learn from you."* Some Mentors described projects that had made an important impact, while others indicated that they had yet to see anything significant come of these efforts.

Importance of Mentors' Staff.

Fellows' experiences are influenced by a variety of people. In particular, the Mentor's direct reports and administrative staff play important roles in the Fellowship experience and can serve as informal mentors. Anna suggested, "*Mentors should set an expectation with their organization that the Administrative Fellow is a Fellow to the organization, not just a Fellow to the vice president."* In reflecting on what he could do better as a Mentor, George said:

> *I have some [staff] who are far less enamored with the program than others, and they're a little resistant and I need to both prepare them and lay out some expectations about this. Why we're doing this, this is what I expect of you in terms of your contribution to*

make sure this is a good experience for this person, and in fact if we do it the right way we should benefit as an organization.

The Mentoring Relationship

When mentoring relationships are assigned the "fit" between a mentor and protégé is uncertain. Mentors felt themselves able to work with a wide variety of potential Fellows, but emphasized the importance of selecting Fellows with the right attitude. This attitude was variously described as positive, assertive, curious, collaborative, and trustworthy. Fellows acknowledged the importance of fit – 84% considered it somewhat or very important – and felt that the Selection Committee did a good job of pairing Mentors and Fellows and that their relationship with their Mentor was generally a positive one. Ninety-five percent of Fellows reported having a good or very good fit with their Mentors.

Not every person is prepared to mentor. Mentors should have an appropriate skill set, be engaged in the process, and be invested in the protégé. Fellows were generally very positive about the level of commitment their Mentors had to the program and to Fellows' professional development. A small proportion, however, felt that their Mentor was not fully engaged. This deficiency was often put in the context of there not being explicit or well-communicated expectations for Mentors. Fellow Sandra said, *"[I would recommend] making sure that the administrator at that level is really, really interested in taking someone on and understands what that word mentor means."* The importance of having a program administrator that they could talk to about difficulties in the mentoring relationship was noted by Fellows, Mentors, and Committee Members.

Program Design

Recruitment and Selection of Fellows.

The selection of Fellows is a competitive process and the AFP represents a significant University investment in the development of a relatively small group of individuals. Selecting Fellows that will take full advantage of the experience is critically important. Mentors wanted Fellows who were self-directed, open-minded, energetic, and collaborative. The importance of seeking people who saw the program as an opportunity rather than as an escape route was particularly noted by several Mentors. Fellows focused on the importance of curiosity, of going into the program as a learner, and of being open to new experiences.

The importance of identifying Fellows at the right point in their career to best benefit from the program and the difficulty of recruiting them was an issue that emerged primarily in discussions with Mentors. Finding the appropriate balance between experience and potential for growth was a balancing point noted by more than one study participant. Some felt that Fellows who already held advanced administrative positions did not gain much from the program. In counterpoint, such Fellows felt that they were uniquely prepared to make the most of the experience because they already had an understanding that less-experienced Fellows lacked.

The majority of Mentors were satisfied with the quality of the Fellows they had worked with and felt that the selection process worked well. However, there were some concerns that the pool of potential candidates was not as deep as it should be and that the quality of Fellows was uneven. Some Mentors expressed uncertainty about the program’s record of identifying the best candidates and acknowledged that they and other University leaders should take more responsibility for identifying and encouraging potential applicants.

Mentor Selection and Preparation.

In discussing the selection of Mentors, both Mentors and Fellows were interested in the possibility of expanding the pool of Mentors. Specifically mentioned was the possibility of including individuals based on their mentoring qualities rather than basing it solely on position. Good Mentors were described as having *"demonstrated leadership,"* and *"the ability to coach."* They were also *"change agents," "well-respected,"* and *"known for giving very development, deliberate, intentional feedback".* Another theme related to Mentor selection was the limitations of the single-mentor model. Both Fellows and Mentors indicated that exposure to multiple mentors and multiple units could enrich the overall experience. Jessica indicated, "*I would love to have had multiple Mentors. I would like to have spent . . . three months with X and three months with Y and three months with Z."*

Sixty-three percent of Fellows reported that their Mentor was well or extremely well prepared to help them make the most of their experience; 32% indicated that their Mentor was somewhat prepared and 5% felt that their Mentor was not at all prepared. Fellows were very positive about the quality of Mentors that have been involved in the program, but both Mentors and Fellows believed that Mentor preparation could be improved. Fellow Mike asked: *"What is the Mentor understanding and do they know what they are supposed to be doing with their mentees to make sure that the mentee gets everything out of it over the year?...I think he didn't quite get all that."*

Most new Mentors had a general understanding of the expectation that Fellows would be shadowing them and that the Fellow should be given entrée into their networks. Mentor Mark said, *"[The program administrator] is very good at explaining what the role is and what the* *expectations are; what the goals of the program are….I thought I was well prepared."* George however, noted that "*I sort of learned by doing it and that was not a good thing."*

Not all Mentors felt that more preparation was necessary and, in general, Mentors believed that they knew how to mentor others. Some Mentors did express a desire for greater preparation and support, and for clearer expectations. Tom, for example, suggested that it might be helpful to have a kickoff meeting with Mentors to talk about ground rules, learning outcomes, and best practices. In reflecting on why this wasn't happening, Tom said, *"there may be a presumption that vice presidents either, 1) know how to do this already or 2) don't have time to [attend another meeting]."* Mentors generally seemed uncertain about what Fellows were told coming into the program and some felt that knowing this would help ensure that everyone in the program was on the same page. Mentors and Fellows felt that selecting Mentors who were new to their positions was detrimental to both the Mentor and the Fellow.

Length of the Program.

The yearlong, full-time commitment of the AFP was a dominant area of discussion in all of the interviews. The program length was established in order to: 1) allow participants to be involved in a unit through a full academic cycle, 2) provide time for trust and communication to develop between the Mentor and Fellow, and 3) provide both breadth and depth for the Fellows. Fellows were not unanimous, but generally saw the length and full-time nature of the program as a strength. Mentors typically were more open to considering either a shorter overall program or less immersive structure, in which Fellows participated in program activities for a certain number of days a month while remaining in their positions. The time commitment was noted as particularly problematic in recruiting high-productivity pre-tenured faculty. Mentor David observed, "*If you are running a lab you can't just say to your grad students, 'Well, I am going to go be an Administrative Fellow. See you next year.'*"

Reservations about the length of the program were often tied to concerns about its lack of structure. Several participants posited that the University should consider either shortening the program or increasing the amount of structure for participants. In arguing for more structure or a shorter program, Mentor Don observed, *"It was, you know, almost a 12-month shadowing experience. . . . Shadowing is interesting, but unless you are really engaged in the work, it has very significant limitations."* Some of the study participants thought that moving away from the full-time commitment and focusing on a more training-oriented model would open the doors to a greater and more diverse range of participants.

Mentors acknowledged the significant time commitment necessary to serve as a Mentor, which may be explain their beliefs that the program should be shortened or that Fellows be given a more concrete task. Shortening the program was also noted as a way to increase the number of participants by allowing more than one cycle of Fellows per year, limit the consequences of poor Mentor-Fellow fit, and encourage participants from other Penn State campuses.

Structure.

The relative absence of required program activities or a curriculum was one of the most talked about components of the AFP. Opinions on the appropriate structure for the program ran from no structure at all to an academy-type structure or curriculum, and appeared unrelated to Fellows' or Mentors' roles (e.g., faculty, administrator, or staff member). Brenda recalled a common frustration among Fellows, *"I found myself with a lot of time on my hands with no constructive purpose to do something with. That was one of the most disappointing points of the Fellowship and one of the most frustrating parts of the Fellowship."* In contrast, Nancy felt that, *"structure means that someone has imposed a structure for you to go through and to learn. And this is the year, for me, free from my teaching, free from my other responsibilities, just to learn."* Like Nancy, many of the Fellows and Mentors felt that the flexibility of the program was one of its key strengths, but others saw it as the program's greatest flaw. In questioning the unstructured nature of the AFP, Mentor Ken observed, *"[The program] shows you how complex things are, the nuances of the Trustees, the President's Council, and all that. That's exposure, but I don't know if it's development."* For the Fellows and Mentors that desired more structure, the nature of that structure varied, but there was general agreement that it should not be too rigid. Fellow Mike said, "*It was good to attend [meetings] and learn from whatever topic was discussed that day, but it would have been nice to have something that would be more . . . like a curriculum."*

Program Outcomes

Fellows were asked a series of survey questions that asked them to judge the efficacy of the AFP in meeting its objectives. Eighty-four percent of Fellows felt that the program met or exceeded their expectations and 79% were satisfied or very satisfied with their ability to meet their personal goals for the program. Fellows were asked to rate the program on each outcome using a six-point scale where: 1 = very ineffective, 2=ineffective, 3=somewhat ineffective, 4=somewhat effective, 5=effective, and 6=very effective. On average, Fellows rated the program as at least somewhat effective, and typically effective or very effective in each objective (Table 3). Fellows generally rated the program higher on providing learning opportunities than on providing opportunities for practice. The highest rating was for the program's ability to enhance understanding of the environment in which University decisions are made.

Table 3. *Fellows' Perceptions of the Effectiveness of the AFP*

Objective	**Mean**	**Standard Deviation**
Enhancing understanding of the environment in which University decisions are made	5.74	0.45
Providing a better understanding of the challenges of higher education administration	5.58	0.61
Increasing awareness of the complexity of issues facing higher education	5.53	0.70
Providing opportunities for learning about the decision-making process	5.47	0.70
Providing opportunities for participation in decision-making processes	4.16	1.50
Providing opportunities for participation in program management	3.89	1.56

Knowledge of the University.

Penn State is one of the largest and most complex institutions of higher education in the world. Although most Fellows came to the program after many years at the University, a primary goal for each was an increased understanding of the different facets of the University and their connections. Reflecting on her program experience, Carol recalled, *"I sought out opportunities for those areas of Penn State that I wanted to know more about."* Mentors likewise felt that the opportunity to increase Fellows' knowledge of the breadth of the University was a foundational function of the AFP. Different Fellows identified different growth areas depending on their mentoring unit, their personal experiences, and their projects, but many mentioned increased understanding of Penn State's complexity as one of the most important things they learned.

Administrative Understanding.

One of the primary goals of the AFP is the development of Fellows' understanding of the roles and skills of administrators, and of the complex and interconnected environment in which decisions are made. When asked about their goals for participating in the program, Mentors were unanimous in this perspective. Scott noted: *"To actually see it as a greater whole is very important, and to be able to go back to their unit and see how that unit participates and contributes to the greater organization is very, very important."*

Based on the experiences of the Fellows interviewed for this project, the program has achieved notable success in this area. Sandra said, "*"I gained a healthier understanding of the complications of running an institution of this size."* Fellows talked extensively about the importance of being exposed to different areas of the University, of considering big-picture questions, and of being exposed to the styles of various University executives. Linda said, *"You can sit back and observe what's successful and what's not"* Similarly, Sharon reflected, "*I know how to be civil, I know how to be an advocate without aggravating people because I learned from the best and I realize that and I am so appreciative of everything that I learned."*

While not an explicit goal of the AFP, an important outcome noted by many of the participants was a greater appreciation for the work, dedication, and commitment of University leaders. Carol observed *"I have gained a better understanding and better appreciation for the many demands placed on the senior administrators. . . . [They] really earn their salaries and they really appear immensely dedicated to their jobs."* Mike reflected on his change of perception, *"[I used to think that] the top, the Old Main building, they don't really think about us. They are just doing whatever they want. And at the end, it was a whole different point of view."*

Professional Advancement.

Not every Fellow enters the program hoping to get a new job afterwards, and most understood that the expectation was that they would return to their original units at the completion of the program. Carol observed, "*Penn State's doing a better job at saying to Fellow* *applicants, this isn't guaranteeing you a promotion, this is guaranteeing you a wonderful opportunity that you need to make the most of*." All of the survey respondents agreed that the Fellowship helps participants (in general) to compete for positions at higher levels of administration (Figure 3) and 89% felt that participation in the program had opened doors to advancement in their own Penn State careers (somewhat agree, agree, or strongly agree). Kim credited the program with having a decisive role in her career progression:

> *I don't want to be overly dramatic, but it changed my life. . . . It totally changed my career path. And I am doing different things that I never thought that I would be doing and I think I have much more, very different and exciting opportunities, that I don't think I would have had before.*

When surveyed about their advancement following the end of the Fellowship, 47% indicated that they had advanced in some way within the first year after completing the Fellowship (Figure 4). Advancement in this context may have been interpreted by respondents to include advancement along the traditional promotional pathways of faculty (e.g., assistant/

Figure 3. Fellows Believe the AFP Helps Fellows Compete for Administrative Positions

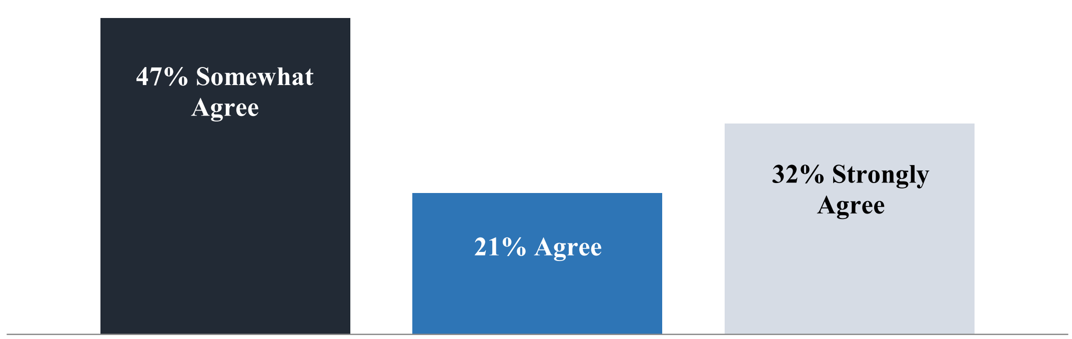
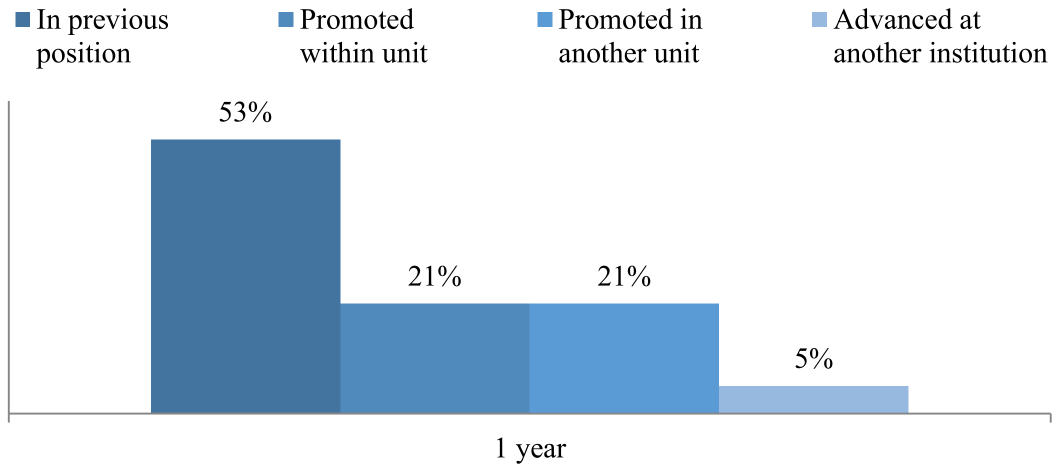


Figure 4. Employment Status One-Year Post-Fellowship



associate/full professor). Based on their survey self-reports, most Fellows (63%) reported advancing in higher education administration after participating in the program. Based on a review of the job titles and career progressions of all Fellows since the program's inception, an estimated 79% advanced in higher education administration. Further, many Fellows who had not changed positions post-Fellowship did take on more responsibility in their existing positions. Among the post-Fellowship job titles of past Fellows are deans, assistant deans, vice provosts, and executive/senior directors. While such evidence supports the efficacy of the AFP, it is not possible to ascertain whether these Fellows – all high-achieving employees with a demonstrated interest in administration – would have advanced regardless of the AFP experience.

Fellows who participated in the Fellowship at least partly as a springboard to a new position but had not advanced, still saw value in the Fellowship experience. Anna captured this feeling when she said, *"My career was not advanced by my Fellowship experience, but my career was enhanced by my Fellowship experience."* Fellows who had advanced in their careers post-Fellowship were generally more positive in their assessment of the AFP.

Mentors frequently mentioned and expressed concern with the expectation of Fellows that they would immediately advance upon completing the Fellowship. Committee Member John noted, "*A lot of times the timing just isn't right. You see that people are really great, but . . . there is just no position. The opportunity just isn't there."* Both Mentors and Fellows, but particularly Mentors, felt that it was important to manage Fellows' expectations in this regard. For example, some Mentors were concerned that Fellows expected a position to be created for them in the mentoring unit or thought that they would not have to compete for open positions. When asked whether having been a Fellow would make someone a more competitive job candidate for a position, one Mentor mused, *"That would be an edge, absolutely. But to say that this is a program that is designed for the next step. . . . I don't know. But I do think it's a great program, provided we're clear about the expectations."* In general, Mentors seemed unclear about the messages that Fellows were getting about the expected outcomes of the program.

Although none of the Mentors described the Fellowship as a way to try out or identify potential new employees, several of them had brought prior Fellows into their units through competitive processes. In talking about this issue, Tom said:

> *I don't think we want to be going around creating positions just so that a Fellow can land in a new spot . . . and at the same time, after they've spent a year kind of following you around as a vice president and so forth, you know about them and they know about you and so it makes a hire easier, because you've basically been interviewing them for a year.*

Better Employees and University Citizens.

As described previously, both Fellows and Mentors spoke to the importance of the AFP in providing an experience that prepares Fellows to advance and also enhances their ability to serve the University in their existing[2] roles. Fellow Jessica recalled:

> *I felt that even if I didn't go anywhere further with it, that I would be able to contribute to the department. I would be able to help my students. I would be able to help the Senate. I would be able to contribute more meaningfully because I knew more about the institution.*

Mentor Scott observed, *"Anyone can aspire to leadership, but it could be leadership because you become a more active member of the unit."* All of the survey respondents agreed or strongly agreed that the Fellowship helped participants to become more effective in their existing positions. Reflecting on returning to her position, Dorothy said:

> *I think others felt that I had knowledge, I had valuable knowledge that they liked knowing I had and it helped them. . . . We would have staff meetings . . . and people would . . . make comments where they really thought the upper administration didn't understand or didn't do things the way that they thought things should be done. I would have the opportunity to say, 'No, it doesn't work like that.' Or, 'No, there's a bigger picture here. You're thinking small, you are thinking just us, but this is how it impacts everybody.' And I think it was my experience as a Fellow and seeing those things, I could bring it to others and then help them to try to see things . . . from another side.*

Some Fellows credited the AFP with opening up other opportunities for professional and personal growth, such as participating on key University committees. They credited these opportunities largely to the knowledge and skills they gained as Fellows, as well as to the

[2] The position they were in upon entering the AFP.

connections they made during the program. Both Mentors and Fellows felt that the alumni Fellows were in a unique position to contribute to the University, no matter what their current role or title, and that they were underutilized post-Fellowship. Committee Member John said:

> *I feel very comfortable that I can go to a former Fellow, and say, 'Look, this is a very sensitive issue. It's going to be very controversial. A lot of confidences need to be in place here. And you've been through this and I think you could do a really outstanding job of either chairing the committee or being an influential member of the committee'. . . . A lot of faculty and staff, absent the Fellowship experience, you couldn't ask them to do this.*

Many Fellows pointed to the networks established during their Fellowship year as one of the most important outcomes of the experience. For example, Brenda said:

> *Getting to know the people, getting to know the structures of the University, how people intersect with one another, who has influence over whom. . . . I now had connections in an area of the University in which I previously had no connection. I could pick up the phone or send an email and people gave me the time of day in a nanosecond. That was the best thing I got out of the Administrative Fellowship.*

The Price of Participation.

Temporarily removing key employees from positions of significant responsibility can leave a void that sending units struggle to fill. For some Fellows, separating from their home unit during the Fellowship year was stressful. For faculty this can mean leaving ongoing research projects, graduate students, and collaborations. For staff, it often means leaving colleagues short-handed. Patricia recalled, "*It's very hard. Because I mean you work for years to build relationships and to put processes in place . . . and then you're just handing it over and praying.*"

In order to fully benefit from the AFP, Fellows are encouraged to separate entirely from their home units for the Fellowship year and many Fellows do not have difficulty doing so. Brenda said, "*I did not have problems separating from my prior role. The office understood what I was attempting to accomplish because it benefited not only me but the office and the University, so it was a win-win-win."* For some, the opportunity to separate was seen as a type of sabbatical, where they were still working, but in a way that rejuvenated them and introduced them to new opportunities and areas for growth. But for other Fellows, conflicting loyalties were a significant source of tension. Fellows generally, but not always, credited the AFP with sending clear messages to the units about the expectations for separation (faculty felt this was less clear than staff), but did not think that this was always realistic. Mike gave an example:

> *I wasn't even done with my Administrative Fellowship, it was done in June, well in summer there was a class, and I needed to teach it. There was no way around it. So I was teaching a class . . . while I was finishing my Administrative Fellowship. . . . So, I was like, 'Here I am again. I am doing two jobs for the next month and a half.' But we have to do it. I mean there was no way around it.*

A number of Fellows spoke of the guilt they felt over leaving their colleagues to pick up the slack in their absence and some Fellows were unwilling to separate because of their concerns about decisions being made in their absence.

For faculty, the Fellowship was often viewed in terms of the trade-off between their administrative interests and progress toward promotion in the faculty ranks. Faculty Fellows are typically tenured associate professors, but that is not the only promotional hurdle that faculty face. Jessica stated:

I knew that taking the Fellowship as a faculty member, meant . . . you were taking a year out of your trajectory toward full professorship. . . . I had to think very carefully about what it meant in terms of my reaching full professorship. So, I decided to go ahead, knowing that it would probably have some implications.

Unit support was an important factor in determining the level of separation difficulty faced by Fellows. Mentor Mark observed, "*If . . . a department head . . . doesn't really understand the purpose of the program or simply isn't as supportive, it can be difficult, awkward."* For Fellows who were encouraged to participate in the program by their supervisors, separation was easier to achieve. Kim described the importance of her supervisor's support: *"She encouraged me and said you need to do this. . . . On my own, I would not have done it because there was just too much going on."*

Using a Mixed-Method Approach for Institutional Research

Qualitative research is an important tool in the toolbox of IR professionals but it is a labor-intensive effort for the typical IR office. In order to reap the benefits of incorporating qualitative research methods into program evaluation, an IR office must take a practical approach to qualitative research. The use of open-ended survey questions or focus groups rather than individual interview can make efficient use of staff and participant time. Likewise, interview or focus group notes, rather than word-for-word transcription, may be sufficient for evaluation purposes. Collaborating with other invested parties and units with complementary expertise (in this case Human Resources and the university's higher education program) can bring additional staff into the evaluation, as well as a broader perspective on the evaluation itself.

Like quantitative analysis, the analysis of qualitative data requires specialized training and should not be undertaken casually. While theory development may emerge from an IR study, the primary focus of an institutional researcher must be on actionable result. Reporting must balance the richness provided by qualitative data and the need to give voice to research participants with the need for brevity, succinctness, and relevance in communicating to University leaders.

Conclusions & Recommendations

Diverse perspectives were provided on how to improve the AFP. While few things were unanimously agreed upon, several themes emerged. These points apply not only to the AFP, but could be applied to other existing mentoring programs and should be considered in the implementation of new programs.

1. Clarify program goals so that expectations are aligned and consistent.
2. Consider providing a structured curriculum focused on specific administrative skillsets.
3. Actively recruit rising stars.
4. Orient and train Mentors.
5. Get buy-in from all members of the mentoring unit.
6. Communicate to Fellows that they must be the drivers of the process.
7. Require Fellows to set goals and monitor progress at regular meetings.
8. Mix and match program models (e.g., short- vs. long-term, full- vs. part-time) to find what works best at your institution.
9. Identify high-priority learning opportunities and make them a formal part of participation.

The AFP is a well-respected program both internally and externally. It has provided a model for similar programs at other institutions around the country (B. Bowen, personal communication, March 21, 2014). In this study, both Mentors and Fellows focused on the importance of growing university leadership from within and for providing unique opportunities for a diverse group of faculty and staff members to learn about their institution's complexities and its leadership. Participants were nearly unanimous in their belief that the University's leadership is too homogenous and that directed efforts were necessary to diversify. Most saw the AFP as continuing to play a role in that effort.

The results of this program evaluation are already being felt in adjustments to Penn State's AFP program. The survey data provides a comprehensive view of what Fellows experience, while the interview and focus group data provided a richness and depth of understanding that would not have been possible with surveys alone. The resources required to conduct mixed-method program evaluations are not insignificant, but the richness of their findings can provide the detailed level of formative or summative assessment needed to justify such resource-intensive programming.

References

Administrative Fellows Review Committee. (2004). *Administrative Fellows Review Committee Final Report and Recommendations*. University Park, PA: The Pennsylvania State University.

Association for Institutional Research. (2014). 2014 Forum Program. Retrieved from http://forum.airweb.org/2014/Documents/2014ForumProgramBookFinal.pdf

Baltodano, J. C., Carlson, S. Jackson, L. W., & Mitchel, W. (2012). Networking to leadership in higher education: National and state-based programs and networks for developing women. *Organizational Behavior and Human Resource Management, 14*(2), 62-78. doi: 10.1177/1523422311428926

Bidwell, M. (2011). Paying more to get less: The effects of external hiring versus internal mobility. *Administrative Science Quarterly, 56*(3), 369-407. doi: 10.1177/0001839211433562

Blake-Beard, S., Murrell, A., & Thomas, D. A. (2006). *Unfinished business: The impact of race on understanding mentoring relationships*. Retrieved from http://www.wininsights.com/wp-content/uploads/2011/10/Race-and-Mentoring.pdf

Carr, K., & Heiden, E. P., Revealing darkness through light: Communicatively managing the dark side of mentoring relationships in organisations (2011). *Papers in Communication Studies*. Paper 26.

Chronus. (2012). *Why corporate mentoring? Five benefits of a workplace mentoring program.* Retrieved from http://chronus.com/wp-content/uploads/2012/11/Five-Benefits-of-a-Workplace-Mentoring-Program.pdf

Killian, C. M., Hukai, D. & McCarty, C E. (2005). Building diversity in the pipeline to corporate leadership. *Journal of Management Development, 24*(2), 155-168. doi: 10.1108/ 02621710510579518

Cox, K. S., & Salsberry, T. (2012). Motivational factors influencing women's decisions to pursue upper-level administrative positions at land grant institutions. *Advancing Women in Leadership, 31,* 1-23.

Dreher, G. F., & Cox, T. H. (1996). Race, gender, and opportunity: A study of compensation attainment and the establishment of mentoring relationships. *Journal of Applied Psychology, 81*(3), 297-308.

Guest, G., MacQueen, K. M. & Namey, E. E. (2012). *Applied Thematic Analysis.* Los Angeles: Sage.

Hewlett, S. A. (2013). *Forget a Mentor, Find a Sponsor: The New Way to Fast-Track Your Career.* Boston: Harvard Business School Publishing Corporation.

Howard, R. D. (Ed.). (2007). *Using Mixed Methods in Institutional Research*. Tallahassee, FL: Association for Institutional Research.

Ibarra, H., Carter, N. M., & Silva, C. (2010, September). Why men still get more promotions than women. *Harvard Business Review, 88*(9), 80-85.

Krathwohl, D. R. (1998). Methods of Educational and Social Science Research: An Integrated Approach. (2nd ed.). Long Grove, IL: Waveland Press, Inc.

Lamoureaux, K. (2013). Five trends in leveraging leadership development to drive a competitive advantage. *Research Bulletin 2013.* Bersin by Deloitte. Retrieved from http://www.ddiworld.com/DDIWorld/media/trendresearch/fivetrendsinleveragingleadershipdevelopment_tr_bersin.pdf?ext=.pdf

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic Inquiry.* Beverly Hills, CA: Sage.

Madsen, S. R. (2008). Women and leadership in higher education: Current realities, challenges, and future directions. *Advances in Developing Human Resources, 14*(2), 131-139. doi: 10.1177/1523422311436299

Morrison, A. M. (1992). The New Leaders: Guidelines on Leadership Diversity in America. New York: Jossey-Bass.

National Center for Education Statistics. (2013). *Digest of Education Statistics.* U.S. Department of Education Sciences. Retrieved from http://nces.ed.gov/programs/digest/current_tables.asp

Office of the Vice Provost for Academic Affairs. (2014). *Administrative Fellows Program*. Retrieved from http://www.psu.edu/vpaa/adminfellows.htm

Peter, K., and Horn, L. (2005). *Gender Differences in Participation and Completion of Undergraduate Education and How They Have Changed Over Time* (NCES 2005–169). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Prescott, B. T., & Bransberger, P. (2012). *Knocking at the College Door: Projections of High School Graduates (eighth edition)*. Boulder, CO: Western Interstate Commission for Higher Education.

Ragins, B. R., & Scandura, T. A. (1999). Burden or blessing? Expected costs and benefits of being a mentor. *Journal of Organizational Behavior, 20*(4), 493-509.

Sandberg, S. (2013). *Lean in: Women, work, and the will to lead.* New York: Alfred A. Knopf.

Touchton, J., McTighe Musil C. & Campbell, K. (2008). *A Measure of Equity: Women's Progress in Higher Education.* Washington, DC: Association of American Colleges and Universities.

Wiley & Sons, Inc. (2015). *New Directions for Institutional Research*. Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1536-075X/issues

Zachary, L. J., & Fischler, L. A. (2011, January). Begin with the end in mind: The Goal-driven mentoring relationship. *Training+Development.* Retrieved from *http://www.centerformentoringexcellence.com/wp-content/uploads/2010/02/The-Goal-Driven-Mentoring-Relationship3.pdf*

FLAGSHIP INSTITUTIONS AND THE STRUGGLE TO COMPETE

Brian R. Lashley, Ph.D.

Assistant Director of Institutional Research

Office of Planning and Institutional Research

Eastern Connecticut State University

Abstract

In this paper, the difficulties Eastern Connecticut State University faces in terms of competing for students with a nearby "flagship" institution are detailed in numerous ways. Multiple sources of information that are typically available to institutional research offices were used to examine the challenges that Eastern and similar institutions face with respect their relative stature versus state flagship universities. The investigation relies heavily on National Survey of Student Engagement (NSSE) data to illuminate actual student experiences in different college settings. Multiple advantages and disadvantages for students and institutions are revealed and discussed based on those data, as well as differences and similarities in terms of cost of attendance.

Eastern Connecticut State University ("Eastern" hereafter) is located in a setting that presents a particular challenge. Its campus is a short drive (approx. 10 minutes) from the campus of a well-known university, the University of Connecticut ("UConn" hereafter). The senior management at Eastern recognizes the difficulty of attracting top donors, top faculty, and top students to its campus when the state's top research institution is only minutes away.

The Office of Planning and Institutional Research ("PIR" herafter) at Eastern has an interest in comparable institutions that affect Eastern's business – application overlap schools, schools attended by applicants who turned down an offer of admission from Eastern, schools that are sources of transfers-in to Eastern, and schools that are destinations for students transferring out of Eastern. UConn is one such school. Additionally, there are other universities with characteristics in common with UConn that attract students away from Eastern. There are other "flagship" institutions that are similar to UConn; they too are colleges that students choose over Eastern, or become transfer destinations after a period of enrollment at Eastern. For these reasons, PIR undertook to study flagship institutions as a whole. Several sources of data are available for that approach.

The study of flagship institutions and how they compare to Eastern began in August of 2014 in Eastern's office of PIR and is ongoing. The following document tells the story of what the data revealed. The research was not intended to affect policy, share a best practice, or test a hypothesis. It was an exploratory effort to discover meaningful facts that might confirm or challenge prevailing views about large, highly visible institutions and smaller regional colleges.

Facets of the Competition for Students

The remainder of the paper focuses on one aspect of Eastern and flagship universities: the competition to attract well-prepared undergraduate students. The discussion touches on common perceptions about different types of colleges, relative costs of attending various colleges, differing levels of resource availability at various colleges, and actual outcomes and experiences students have at certain types of colleges.

Common Perceptions About Large Universities

Large universities often play a central role in local or regional culture. Even those who don't attend them are familiar with them through sports, local news, conversations, inclusion in movies or TV shows, etc. As a result, people have perceptions about those universities regardless of any direct experience.

Football, Fun, Social Experiences. College sports are a major part of public connection with higher education today. Sports television broadcasts not only feature sports competitions, but also scenes of non-athlete students at the competing universities enjoying a fun day or evening. Moreover, those scenes focus on groups of people, usually including both genders. These images may affect the perceptions viewers have about the social, developmental, and pleasurable opportunities at the schools whose sports teams are being broadcast. Additionally, aside from television and sports, it is also generally known that large colleges feature attractive social organizations such as the Greek system and other opportunities to mingle with fellow students.

Affiliation Opportunities. One commonality of Eastern and flagship universities is a focus on the "traditional" freshman – a recent high school graduate, approximately 17-19 years old who is enrolling in college for the first time with the intent of earning a baccalaureate degree. Such students are in a certain phase of personal development – a time of life characterized by identity formation, of developing a personality and self-image. Such concerns commonly lead people to affiliate with something larger than themselves – something that is large, well-known, well though of, and perhaps powerful. Large universities often carry a certain prestige, or "big-time-ness" that can be very attractive to a young person making their way in the world.

Academic Programs. Larger institutions are able to offer larger arrays of academic major programs. Smaller colleges are often unable to offer highly-sought but expensive (to the institution) majors such as engineering or nursing. The present author attended a flagship institution as an undergraduate and witnessed an acquaintance who majored in an uncommon major called Media Arts and went on to a successful career.

Perceptions About the Employability and/or Salary of Graduates. The use of higher education data in forming decisions about what college to attend has not become widespread, at least for typical high school students and their parents. Hearsay and opinion regarding the quality of an institution is generally known to affect choice of college and may be related to the affiliation needs described above. A common belief is that the name of the college attended affects career success; for example, that "Rutgers" makes a job résumé look much better than "Ramapo," "Texas" looks better than "Midwestern State," and "Washington" looks better than "Evergreen State."

Eastern's Concern About Losing Students to Flagship Institutions

Eastern president Elsa Núñez showed her awareness about the state of the competition in her university address on her Aug. 2015 powerpoint; her address expressed concern about retaining Eastern's undergraduate students and reducing transfers out. In particular, "the lure of a bigger school" was featured prominently in her speech, with "signs of competition" for students in terms of UConn's goal of adding 6500 students and out-of-state colleges advertising in Connecticut newspapers and television channels.

Cost Competition

In addition to the flagships' attractiveness to young people, they can be surprisingly competitive on cost of attendance. Table 1 shows that despite higher tution and fee charges, in terms of net price, some prominent northeastern flagship institutions are in the same cost range as Eastern.

Table 1

Tuition & Fees, Net Price, and Institutional Grant Funds Awarded

Institution	In-State Tuition + Fees (2013-14)	Net Price (2012-13)	Inst. Grant Total $
U of New Hampshire	16,496	21,545	19.5m
U of Vermont	15,718	15,793	25.5m
Rutgers Univ.	13,499	16,040	17.3m
U of Massachusetts	13,443	19,120	15.9m
U of Rhode Island	12,450	17,090	20.5m
U of Connecticut	12,022	18,411	14.4m
Eastern Conn. St.	9376	16,333	1.9m

The tuition and fees and net price figures are from the federal College Affordability and Transparency Center (2015); the institutional grant totals are from each school's IPEDS Financial Aid report, Part C, line 04 (NCES, 2015).

The difficulty for Eastern that is portrayed in the table is the fact that, while its "sticker" price is lowest, its net price is only the third lowest of the seven; in fact, it is not much different from the other six. The institutional grant totals show that Eastern is not able to lower its net price through its own financial aid. These institutional grant totals include both merit aid and need-based aid. Thus, Eastern lacks a certain financial resource that is more plentiful for flagship institutions – meritorious and/or need-based financial aid for undergraduate students.

Transfers Out

National Student Clearinghouse (NSC) data allow for the tracking of students who leave a given institution. Also, the NSC Research Center offers each member institution a Cohort Completions Report that shows how a given Fall cohort is faring six years after enrollment at the member institution. The results are compared to a national average. Table 2 shows Eastern's results for the 2008 cohort (from the StudentTracker Postsecondary Completions report, 2015, tables 2A and 2B).

Table 2

Six-Year Outcomes for Students Who Started at Eastern

	Eastern	Four-Year Publics
Completion at Same Institution	46.1%	49.8%
Completion at Different Institution: 4-Yr	**16.6%**	9.6%
Completion at Different Institution: 2-Yr	3.8%	3.5%

The fall 2008 cohort at Eastern had a similar completion rate at the "Same Institution" and at different 2-year institutions as 4-year public colleges across the United States. However, the completion rate at different 4-year institutions is considerably higher for Eastern; 7% more of Eastern's entering freshman students completed at a different 4-year college than did the same cohort, on average, at all public 4-year colleges. Thus, Eastern does have a relatively difficult challenge with respect to retaining capable college students. One out of six members of its fall 2008 cohort have earned a degree at some other 4-year college.

Educational Differences Between Eastern and Flagships

For an aspiring college student, it would be appropriate to ask what the difference is between the education afforded by Eastern and that offered by the universities of New Hampshire, Vermont, Alabama, Texas, Arkansas, California, Delaware, etc. To address this question, National Survey of Student Engagment (NSSE) data were analyzed.

Nationally, 1.4 million college students were invited to take NSSE in 2015; 315,815 of them responded by taking the survey. At Eastern, 2076 students were invited to participate, and 676 (32.6%) responded by taking NSSE.

Eastern has participated in NSSE every year since 2006. It should be noted that at Eastern the average GPA of NSSE respondents has generally averaged around 3.20, and that the sample is around 70% female in the 10 years that Eastern has participated in NSSE. The average GPA of Eastern students in general is closer to 2.80, and the student body is only about 54% female.

Methodology

NSSE aims its survey at two levels of college students: freshmen and seniors. NSSE also allows participating institutions to create comparison groups of institutions; the NSSE manager at a participating institution can see a list of all institutions participating in NSSE and use the list to create a comparison group. The aggregated frequencies of that comparison group(s) appear on the participant's NSSE Institutional Report. Eastern's office of PIR took advantage of this opportunity, and created a flagship institution comparison group. The present paper compares Eastern's survey results to the collective results of the flagship institutions (N = 22 universities; see list in Appendix A).

Eastern's NSSE 2015 results were compared to the flagships' in terms of "percent favorable" responses to each survey item. For example, one NSSE item asks how much the student's coursework has emphasized the application of facts, theories, or methods to practical problems; the student's response options are Not at All, Somewhat, Quite a Bit, or Very Much. The "percent favorable" in this case would be total number of students responding Quite a Bit or Very Much (ie, favorably) divided by the total number of students who answered the question (excluding missing data). This percent favorable methodology was applied to each NSSE item, separately for freshmen and seniors, for Eastern and for the flagship comparison group. The difference between the percents favorable were calculated by simple subtraction. Finally, the largest five positive differences and the largest five negative differences were used to create a list of "Highest performing items compared to flagships" and "Lowest performing items compared to flagships" separately for freshmen and seniors. The remainder of the present paper focuses on these highest and lowest performing items.

The same analysis was performed in 2014 (the results were not published). In many cases the same items appeared in both 2014 and 2015 highest and lowest items lists. Those items are flagged in tables 3–8, 10, and 11 with an asterisk.

Eastern Freshmen

Table 3

Highest Performing Items Versus Flagships

Item	Eastern Advantage
Faculty provided prompt and detailed feedback on tests or assignments	+13%
Faculty provided feedback on a draft or work in progress	+14%
Discussed course topics or concepts with faculty member outside of class	+12%
Discussed your academic performance with a faculty member*	+12%
Gave a course presentation*	+19%

These five NSSE items – in which Eastern freshman responded with a higher percent favorable – all depend on 1) having small enough classes for individual attention to students, and 2) that the classes be taught by the faculty rather than graduate assistants. Discussing course topics and the student's academic performance with faculty are important face-to-face interactions that enhance the student experience (NSSE, 2015). Giving a course presentation is an item that appears frequently in the remainder of the present paper. The two items with asterisks are items that were also on this list in the analysis of 2014 NSSE data; thus, it appears that they may be stable, ongoing advantages for Eastern.

Table 4

Lowest Performing Items Versus Flagships

Item	Eastern Disadvantage
If you could start over again, would you go to the same institution you are now attending?*	–18%
Hours per week preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing and other academic activities)*	–19%
How would you evaluate your entire educational experience at this institution?	–8%
Indicate the quality of your interactions with academic advisors at your institution	–8%
Plan to participate in a study abroad program	–11%

The results for the first item on the list suggest that Eastern freshmen are considerably more likely to question or regret their decision to come to Eastern. There are numerous reasons why they might feel such doubts; years of freshman retention research at Eastern have shown that there is not one specific reason Eastern freshmen often do not return for their second academic year. Rather, the reasons are many and varied. In light of the competition with flagship institutions, many positive attributes of flagships were noted above – in many cases they are attributes that Eastern cannot match, and could be reasons that flagship freshmen report a higher level of confidence that they made a good choice.

There is also a considerable difference between Eastern freshmen and flagship freshmen in terms of time spent on academic work. The percent favorable for this item was the percent who reported eleven hours or more per week studying and preparing for class. Clearly, flagship freshmen are self-reporting more time on academic tasks than Eastern freshmen.

The first two items are repeated from last year's analysis, so the evidence grows stronger that Eastern freshmen are unsure of their decision to attend Eastern and don't study as hard as flagship students.

Fewer Eastern freshmen indicated that they would participate in study abroad programs than did flagship freshmen. This item appears again in tables 6, 8, and 11, and is detailed further in Table 12.

Eastern Seniors

Table 5

Highest Performing Items Versus Flagships

Item	Eastern Advantage
Working for pay off campus	+17%
Reviewed your notes after class	+13%
Discussed your academic performance with a faculty member*	+17%
Gave a course presentation*	+15%
Asked questions or contributed to course discussions in other ways*	+15%

Again there is evidence of small classes offering the opportunity to have face-to-face interaction with faculty, by discussing academic performance, giving course presentations, or asking questions and being part of course discussions. The last three items are dependent on having small enough classes for individualized attention and conversation; and that there be feedback from a faculty member rather than a graduate assistant. These same three items are also repeated from 2014.

The 'Working for pay off campus' item could be a positive boon or a burden to the student. On one hand, Eastern students are having early workplace success, even if the job they are doing does not require college-level skill. On the other hand, ideally, Eastern students would focus their attention purely on their campus life and have enough money to pay for college without working.

Table 6

Lowest Performing Items Versus Flagships

Item	Eastern Disadvantage
Attending campus activities and events (performing arts, athletic events, etc.)	–10%
Providing support for your overall well-being (recreation, health care, counseling, etc.)	–8%
Hold a formal leadership role in a student organization or group*	–11%
Participate in a study abroad program	–8%
Conversations with people with religious beliefs other than your own	–7%

A surprising item on this list is the third one, regarding the taking of a leadership role in some campus club or other organization. It is not clear why Eastern students would be less likely than their flagship peers to take an opportunity to lead when one considers that the campus is smaller at Eastern. We have seen in this paper that 1) Eastern does not have as much grant and scholarship money available as flagships (Table 1) and 2) Eastern seniors are more likely to hold an off-campus job while still enrolled (Table 5). Thus, while it is conjecture, it could be that they are too busy or mentally occupied with financial concerns to take on the additional responsibility of leadership in a club, student government, or other initiative.

Colleges and Universities Similar to Eastern

Another question this paper will address is whether the advantages and disadvantages described above are particular to Eastern, or whether other institutions – similar to Eastern in important respects – have a similar pattern of results. To investigate this question, another set of NSSE comparison institutions was developed and utilized. The Council of Public Liberal Arts Colleges (COPLAC) is a group of similar institutions, and Eastern is a member. The most basic commonalities for all COPLAC schools are a liberal arts mission and public control. Other aspects that are similar amongst these colleges are 1) similar locations: rural or small town, 2) finances, 3) size in terms of enrollment, 4) preparedness of admitted freshmen, and 5) faculty salaries.

The COPLAC group consists of all COPLAC schools that participated in NSSE 2015 (N = 24 institutions; see list in Appendix B). The previous highest/lowest performing items analysis was repeated with the COPLAC group substituted for Eastern in comparisons to the flagship institutions. The remaining tables in this paper focus on comparisons between COPLAC and flagships, with Eastern held out of the comparisons.

COPLAC Freshmen

Table 7

Highest Performing Items Versus Flagships

Item	COPLAC Advantage
Of the time you spend preparing for class in a typical 7-day week, about how much is on assigned reading?	+12%
Instructors provided prompt and detailed feedback on tests or completed assignments*	+8%
Instructors provided feedback on a draft or work in progress*	+9%
Gave a course presentation*	+13%
Attended an art exhibit, play or other arts performanc (dance, music, etc.)*	+10%

Like Eastern freshmen, COPLAC freshmen are more likely to give a presentation for a class in the first year of college. They are also more likely to receive feedback from faculty that freshmen at flagship schools. The second, third, and fourth items on this list overlap with the Eastern freshmen's advantage list. Also, all except the first item are repeats on this list from the 2014 analysis.

Table 8

Lowest Performing Items Versus Flagships

Item	COPLAC Disadvantage
Participate in a learning community or some other formal program where groups of students take two or more classes together	–7%
Preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities)*	–7%
Plan to participate in a study abroad program*	–9%
Hold a formal leadership role in a student organization or group*	–11%
Asked another student to help you understand course material*	–7%

COPLAC freshmen, like Eastern freshmen, do not appear to be spending as much time on academic work as freshmen in flagship schools. Nor do they appear to have plans for study abroad as much as flagship freshmen. The second and third items on this list overlap with Eastern's freshman advantage list. It appears that, although large colleges may be associated with football, parties, Greek systems, etc., the average student studies more at these institutions. They are also more likely to study in another country.

Table 9

Comparison of Eastern, COPLAC, and Flagships on Hours Per Week Preparing for Class

Percentage of NSSE Respondents Reporting More Than 10 Hours Per Week Preparing for Class

Level	Eastern	COPLAC	Flagships
Freshman	47%	58%	66%
Senior	57%	62%	63%

Table 9 focuses in on one NSSE item in particular, the item that asks students to estimate their weekly hours spent on academic work. That item has appeared on each freshman disadvantage list, and is one of the more revealing findings in this paper. Table 9 displays the percentage of both freshmen and seniors who report spending at least 11 hours per week on studying, papers, analyses, etc. related to coursework. It shows that at both the freshman and senior levels, flagship institutions have the highest percentage. Especially concerning to the present author is the large difference between Eastern freshmen and flagship freshmen. The difference between these two is not as large at the senior level, although flagship institutions still have the most reported time on academics.

COPLAC Seniors

Table 10

Highest Performing Items Versus Flagships

Item	COPLAC Advantage
Completed a culminating senior experience (capstone course, senior project or thesis, comprehensive exam, portfolio, etc.)*	+17%
Discussed your academic performance with a faculty member*	+10%
Of the time you spend preparing for class in a typical 7-day week, about how much is on assigned reading?	+14%
Interactions with administrative staff and offices (registrar, financial aid, etc.)	+11%
Instructors provided feedback on a draft or work in progress	+11%

COPLAC seniors are more likely to engage in behaviors that require close attention from faculty, although the item on discussing academic performance with faculty is the only one on the list that overlaps with Eastern's seniors' highest-performing item list. The culminating senior experience item will be discussed further below.

Table 11

Lowest Performing Items Versus Flagships

Item	COPLAC Disadvantage
Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)*	–4%
Analyzing numerical and statistical information*	–8%
Attending campus activities and events (performing arts, athletic events, etc.)*	–5%
Participated in a study abroad program*	–6%
People of a race or ethnicity other than your own	–6%

COPLAC seniors do not seem to have as much experience with quantitative reasoning as flagship seniors, based in the first two items. These two items do not overlap with Eastern's 2015 seniors' lowest-performing items list, but they do overlap with the 2014 list. These items are repeated from 2014 for both the COPLAC-flagship senior comparison and the Eastern-flagship senior comparison.

One item that does overlap with the analogous Eastern table is the 'attending campus activities' item. Since the first example mentioned in parentheses in the wording is "performing arts," it will be of interest to see if the ratings change for Eastern after its new Fine Arts Center opens in the winter of 2016.

The final table in the present paper focuses on a NSSE item that may be of particular interest to COPLAC institutions, and possibly all institutions. The item asks respondents to indicate their participation in certain "high-impact practices" identified by scholars (NSSE, 2015). The table reflects percentages of seniors reporting that they did experience the given practice, and includes Eastern, COPLAC, and flagship institutions.

Table 12

Seniors' Participation in High-Impact Practices

High-Impact Practice	Eastern	COPLAC	Flagships
Internship	56%	54%	58%
Learning Community	29%	24%	26%
Study Abroad	13%	16%	21%
Research with Faculty	34%	32%	30%
Culminating Senior Experience	51%	60%	42%
Service-Learning	61%	66%	52%

Note: Percentages reflect students who reported they have participated in the listed activities. Student responses to *service-learning* indicate that *at least some* of their courses included a service-learning experience.

Many of the "lowest-performing items" tables have featured the study abroad item. Flagships are clearly more able to get students into that particular type of learning opportunity. For the other practices listed on the table, the variation among the three comparison groups is limited. For example, the participation rate in internships only ranges from 54% to 58%. The rates for culminating senior experiences may be noteworthy though, as considerably more COPLAC seniors reported participation that Eastern flagship seniors.

Conclusions

This paper has told the story of a medium-size public regional college's struggles in light of its close proximity to a well-known flagship institution. The focal point of the story is undergraduate students – attracting quality applicants, getting them to enroll, and getting them to stay and earn their degree at Eastern. Perhaps the key to this whole paper is the fact that the NSSE Item "If you could start over again, would you go to the same institution you are now attending?" is a low-performing item for Eastern freshmen, but not a low-performing item for Eastern seniors or COPLAC students at either level. Eastern has never been able to reach 80% freshman 1-year retention, despite years of effort and programmatic improvements aimed at achieiving it. Yet, 4- and 6-year graduations have climbed noticeably in recent years. It may be that many Eastern freshmen ignore the advantages of face-to-face interactions with faculty and maintain a mindset that "somewhere else is better." Those who see the advantages of these interactions may be the ones most likely to still be at Eastern for their senior year.

One low-performing item that both Eastern and COPLAC feature when compared to flagships is the amount of time spent studying or doing other academic work. There could be many explanations for the difference. For example, since flagships and their campuses and activities are so attractive, the most academically-focused high school students may constitute a majority of the admitted freshmen at these institutions. Their cognitive level then allows the faculty to teach more complex and demanding material, requiring more time studying; the combination of student body academic capabilities with faculty academic demands could then feed off of each other to create a culture of high expectation and more time on academic tasks.

It is possible that the feeling of school pride amongst current students and alumni is higher with flagship institutions than at Eastern. That could indirectly affect retention, survey ratings, and even cost of attendance. Table 1 revealed that, despite higher tuition and fee charges, flagship universities are as affordable as Eastern in terms of net price. Table 1 also implied that this cost competition may be driven by flagships' greater resources in terms of grant and scholarship money for undergraduate students. Although the present paper does not go as far as identifying the source of flagships' financial aid resources, one may conjecture that the main source is these universities' foundations. These foundations build resources through fundraising, which is generally more successful when alumni and businesses have a positive view of the university.

References

College Affordability and Transparency Center (2015). *Public, 4-year or above with Highest Net Prices* [data file]. Retrieved from http://demos.inovas.net/collegecost/catc/default.aspx

National Center for Education Statistics (2015). *IPEDS Data Center*. Available from https://nces.ed.gov/ipeds/datacenter/Facsimile.aspx

National Student Clearninghouse Research Center (2015). *StudentTracker Postsecondary Completions*. Provided for Eastern Connecticut State University.

National Survey of Student Engagement (2015). *Institutional Report 2015.* Provided for Eastern Connecticut State University.

Appendix A

Flagship Institution Comparison Group

Central Connecticut State University (New Britain, CT)*

Rutgers University-New Brunswick/Piscataway (New Brunswick, NJ)

Southern Connecticut State University (New Haven, CT)*

University at Buffalo, State University of New York (Buffalo, NY)

University of Delaware (Newark, DE)

University of Georgia (Athens, GA)*

University of Hawaiʻi at Manoa (Honolulu, HI)

University of Idaho (Moscow, ID)

University of Illinois at Urbana-Champaign (Urbana, IL)

University of Kentucky (Lexington, KY)

University of Maryland (College Park, MD)

University of Massachusetts Amherst (Amherst, MA)

University of Mississippi (University, MS)

University of Missouri-Columbia (Columbia, MO)

University of New Hampshire (Durham, NH)

University of North Dakota (Grand Forks, ND)

University of Oregon (Eugene, OR)

University of South Carolina Columbia (Columbia, SC)

University of South Dakota (Vermillion, SD)

University of Vermont (Burlington, VT)

University of Wisconsin-Madison (Madison, WI)

University of Wyoming (Laramie, WY)

* Not a flagship institution; included by mistake

Appendix B

COPLAC Comparison Group

Evergreen State College, The (Olympia, WA)

Fort Lewis College (Durango, CO)

Georgia College & State University (Milledgeville, GA)

Henderson State University (Arkadelphia, AR)

Keene State College (Keene, NH)

Mansfield University of Pennsylvania (Mansfield, PA)

Massachusetts College of Liberal Arts (North Adams, MA)

Midwestern State University (Wichita Falls, TX)

Ramapo College of New Jersey (Mahwah, NJ)

Sonoma State University (Rohnert Park, CA)

Southern Oregon University (Ashland, OR)

St. Mary's College of Maryland (Saint Mary's City, MD)

State University of New York at Geneseo, The (Geneseo, NY)

Truman State University (Kirksville, MO)

University of Illinois Springfield (Springfield, IL)

University of Maine at Farmington (Farmington, ME)

University of Mary Washington (Fredericksburg, VA)

University of Minnesota, Morris (Morris, MN)

University of Montevallo (Montevallo, AL)

University of North Carolina at Asheville (Asheville, NC)

University of Science and Arts of Oklahoma (Chickasha, OK)

University of South Carolina Aiken (Aiken, SC)

University of Virginia's College at Wise, The (Wise, VA)

University of Wisconsin-Superior (Superior, WI)

USING DATA MINING TO PREDICT FRESHMEN OUTCOMES

Nora Galambos, PhD
Senior Data Scientist
Office of Institutional Research, Planning & Effectiveness
Stony Brook University

Abstract

Data mining is used to develop models for the early prediction of freshmen GPA. Since student engagement has long been associated with student success, the use of service utilization and transactional data is examined along with more traditional student factors. Factors entered into the data mining models include advising visits, freshmen course-taking activity, interactions with the college learning management system, and college activity participation, along with SAT scores, high school GPA, demographics, and financial aid. In models predicting first semester freshmen GPA, factors associated with students' interactions with the campus environment were stronger predictors than SAT scores.

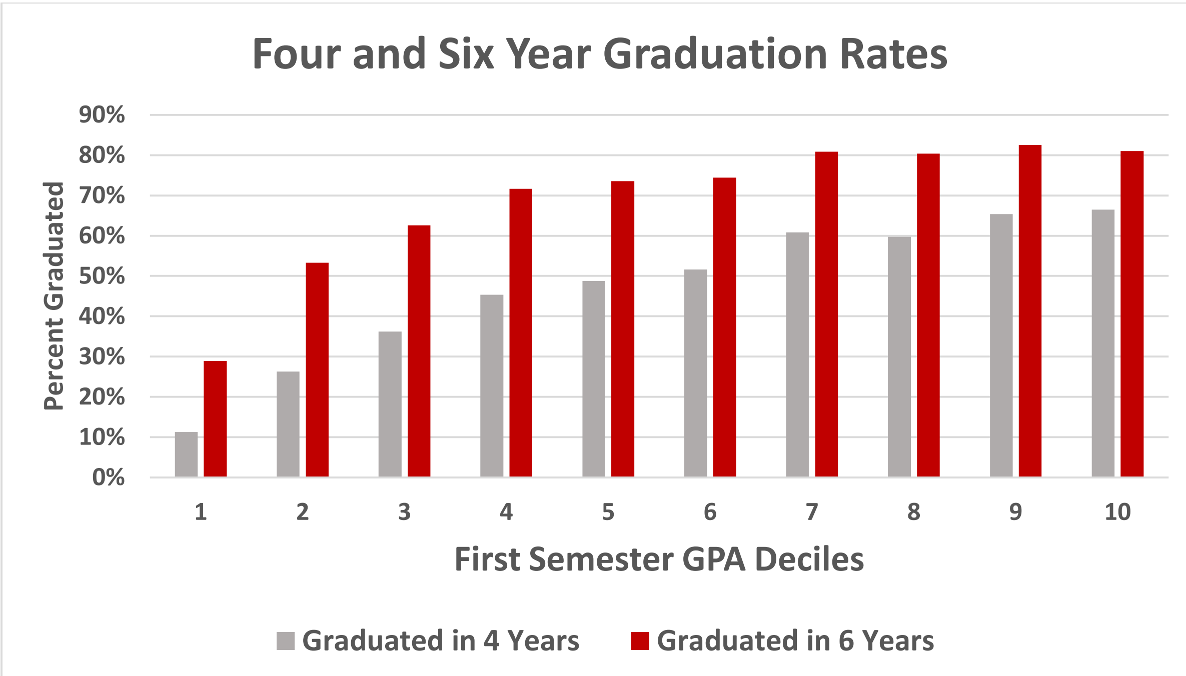
Introduction

The goal is to develop a model to predict at risk first-time full-time freshmen as early as possible in their college careers in order to assist them with interventions. Traditional methods of logistic and linear regression are often good at identifying factors significantly associated with an outcome, but are not always able to make accurate predictions. Linear and logistic regression have one set of predictors to model the outcomes of all of the students in the data and do not assign separate sets of predictors to students having very different characteristics. For example, first-time freshmen entering college with high SAT scores may have very different retention and college GPA predictors than those entering with a low high school GPA and low SAT scores. Inevitably, when using any model, some students will be incorrectly assigned, with some students miss-identified as being at risk or students at risk being not being identified as such by the model. There is an allocation trade-off when resources are expended on students not really in need of interventions or when students who would potentially benefit from interventions do not receive them. Methods capable of more accurate predictions will result in more effective utilization of resources, and higher retention and graduation rates. For that reason the decision was made to explore data mining, because it offers a variety of methods for utilizing different types of data, there are few assumptions to satisfy relative to traditional hypothesis driven methods, and it is able to handle a great volume of data with hundreds of predictors.

At our institution poor academic performance by first-time full-time freshmen in the first semester has a negative impact on graduation and retention outcomes. Figure 1 illustrates that only 11% of students in the lowest GPA decile graduate in four years, and less than 29% of students in that group graduate in six years. For the second decile the four year rate increases to 26% and the six year rate improves to 53%. Those rates, though higher, are still very low relative to the top half of the freshmen class.

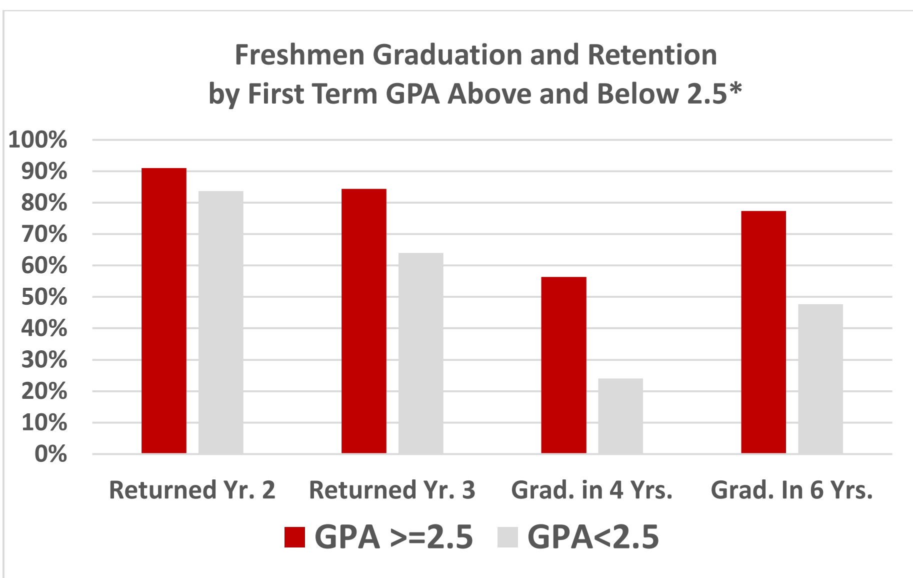
Approximately 30% of first-time full-time freshmen received a GPA below 2.5 in their first semester (Figure 2). Almost 84% of those students returned in year two, however by the next year the retention rate had dropped substantially with only 64% returning for year three and only 48% graduating in six years. In contrast over 77% of students receiving a GPA of 2.5 or greater in their first semester graduated in six years.

Figure 1. Four and Six Year Graduation Rates of First-Time Full-Time Freshmen by GPA Deciles*



*The fall freshmen cohorts of 2006 through 2008 were combined.

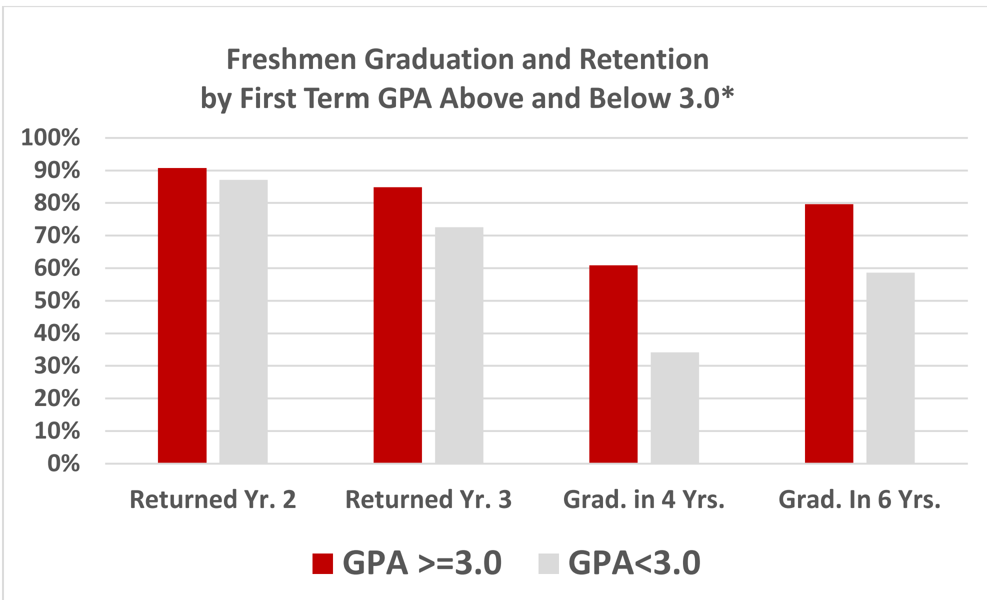
Figure 2. Comparison of Graduation and Retention Rates of First-time Full-time Freshmen by First Semester GPA Above and Below 2.5.



*The fall freshmen cohorts of 2006 through 2008 were combined.

Even when evaluating results for students above and below 3.0 the differences are dramatic (Figure 3). Only 34% of students with a first semester GPA below 3.0 (approximately the median) graduated in four years, which is almost 27 percentage points lower than students above the median.

Figure 3. Comparison of Graduation and Retention Rates of First-time Full-time Freshmen by First Semester GPA Above and Below 3.0.



*The fall freshmen cohorts of 2006 through 2008 were combined.

Given these results we see that it would greatly benefit at risk students if they could be identified as early as possible. In order for the programs to be cost effective and, more importantly, a good match for the needs of the students, the model must be able to make very accurate predictions. The difficulty of this task lies in the fact that there are not many university-level academic measures available on or before the middle of the first semester of the freshmen year. For that reason we have explored the development of a data mining model that combines transactional data such as learning management system (LMS) logins and service utilization such as advising and tutoring center visits with other more traditional measures in an attempt to identify at risk students *before* any grades appear on their transcripts.

Literature Review

The study has cast a wide net in terms of assembling a variety of data for use in studying academic, social, and economic factors to determine elevated risk of a low GPA, which can translate to increased risk of early attrition or longer time to degree. Consistent with the retention study of Tinto (1987), we evaluate many types of data representing students' interactions with their campus environment to determine if higher levels of campus engagement are predictive of improved freshmen outcomes. These measures of engagement include interactions with the learning management system, intramural sports and fitness class participation, and academic advising and tutoring center visits. It appears that students who are identified to be at risk in their first term and remain at the institution, continue to be at risk, with greater numbers leaving in the subsequent term (Singell and Waddell 2010). This is consistent with the results at our institution which are presented in Figures 1, 2, and 3. Methods capable of more accurate predictions will result in more effective utilization of campus resources, and higher retention and graduation rates. Course-taking behavior is also important, particularly math readiness. Herzog (2005) found math readiness to be "more important than aid in explaining freshmen dropout and transfer-out during both first and second semesters." Herzog also focused on both merit and need-based aid and the role that interaction of aid and academic preparedness plays in student retention. Living within a 60 mile radius of the institution, the percent of students at a high school who take the SAT, along with the percentage at the high school receiving free lunches was explored by Johnson (2008) underlining the need to examine the role of the secondary school and socio-economic factors in developing a model. Persistence increases among students closer to the institution and not surprisingly, decreases among those who were from schools having a high percentage of students receiving free school lunches. The role of differing stop-out patterns exhibited by grant, work-study, and loan recipients (Johnson 2010) demonstrated that grants have the highest positive effect on persistence, but its effect decreases more than that of loans after controlling for other factors. Resource utilization was studied (Robbins et al. 2009) using a tracking system. Services and resources were grouped into academic services, recreational resources, social measures and advising sessions, with all but social measures demonstrating positive associations with GPA even after controlling for other demographic and risk factors. These papers have demonstrated that researchers are examining a range of factors in studying and modeling risk. This research underlines that fact that student success is the result of complex interactions between student engagement, academic service utilization, financial metrics, and demographics, which are combined with student academic characteristics that include high school GPA and SAT scores. Data mining is ideal for developing a model with a large diverse number of predictors.

Data Sources

An attempt was made to include as many types of data as possible, so learning management system logins, not previously explored by our institution were included. Building the dataset began with the traditional measures such as demographics (gender, ethnicity, and geographic area of residence when admitted), to which were added high school GPA and SAT scores. In order to control for high school GPA, the average SAT scores of the high schools were incorporated. Because we are modeling the freshmen GPA at the mid-semester point, in terms of college academic characteristics we only have available the fall semester courses in which the students are enrolled, the area the major, whether a major has been declared, and how many college credits were accepted by the institution upon admission. The number of AP credits received was also captured, with those credits separated into STEM and non-STEM totals.

To explore the effect of high failure rate courses on student outcomes, courses with enrollments of 70 or more students having 10% or more D, F, or W grades were identified and categorized as STEM or non-STEM courses. The total number of high DFW-rate courses, and the highest DFW rate for each student (by STEM indicator) was included in the model. The percentage of freshmen in each DFW course was also tabulated and that percentage for the corresponding course was additionally added. The rationale for examining the percentage of freshmen in these difficult courses is that if the courses are populated by large numbers of upper level students, it may make the course even more difficult for freshmen who are less experienced.

Since student engagement has long been associated with student success, the use of service and academic utilization data was included to determine if it resulted in improved models. Student interactions with the university's learning management system, academic advising, tutoring center visits, intramural sports, and fitness classes, have been incorporated in the analysis to evaluate the association of GPA with students' engagement in the university environment.

Much of the data pertaining to interactions with student services and learning management system logins has not been stored long term. In fact the LMS login data was not available for any fall semester prior to fall 2014. As a result, part of the data mining process has included the initial collecting, saving, and storing of the data. Programs are being developed to automate the formatting and aggregation of the transactional data so it can easily be merged with student records and utilized in the data mining process. For modeling use of the LMS logins, only one login per course per hour was counted, so an individual course can have at most 24 logins per day. This eliminated multiple logins that occurred just few minutes and sometimes a few seconds apart. Further, the courses were categorized as STEM or non-STEM. Next the STEM and non-STEM logins were totaled for week 1 and separately for weeks 2 through 6. Finally the STEM and non-STEM logins were divided by their respective STEM and non-STEM course totals to obtain per-course login rates.

Financial aid data was also assembled. The measures that were captured are the expected family contribution, adjusted gross income (AGI), types and amounts of disbursed aid (athletics aid, loans, grants, scholarships, and work-study). Pell Grants and the Tuition Assistance Program (TAP) recipients were also added to the model.

Because the data mining initiative is new and many data sources are being collected and explored for the first time, research and evaluation of the methods for summarizing and using the data in the model is ongoing. The expectation is that additional data sources will be added. A detailed list of the data elements can be found in the appendix.

Methodology

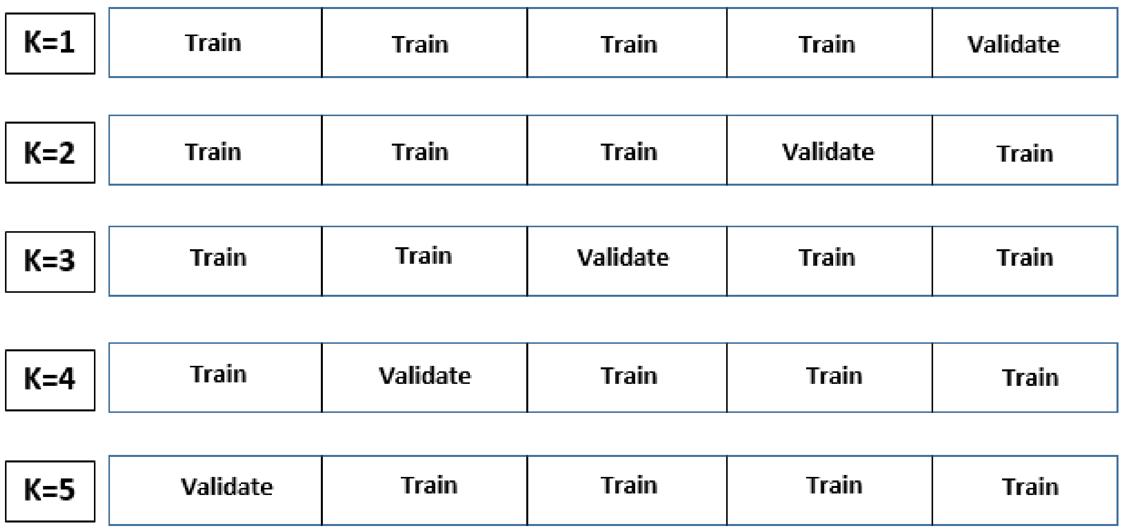
Different models were compared to find the ones that provide the most accurate prediction of the first semester GPA with the lowest average squared errors (ASE)[1] . In developing data mining models it is advisable to partition the data into training and validation

[1] ASE = SSE/N or ASE = (Sum of Squared Errors)/N

sets. The training set is used for model development, then the model is run on the validation set to check its accuracy and calculate the prediction error. It is also important to avoid developing an overly complex model, overfitting. If the model is too complex it can be influenced by random noise, and if there are outliers an overly complex model may be fit to them. Unfortunately, when using such a model on new data its ability to accurately predict the outcomes will be diminished. One way of detecting overfitting is to compare the ASE of the training and validation data. A large increase in the ASE when running the model on the validation data may be a sign of overfitting. However, with less than 3,000 subjects and over 50 variables to predict the GPA's of the bottom 20% of the class, setting aside 40% of the data as is typical for a validation set, is not practical because it would not leave enough of the lower GPA students for building the model. As an alternative, k-fold cross validation was used. It works with limited amounts of data, and its initial steps are similar to traditional analysis. The entire dataset is used to choose the predictors and the error is estimated by averaging the error of the *k* test samples. In subsequent years, when more than one semester of LMS data has been collected, the easier to implement training-validation-partitioning method can be used.

To implement k-fold cross validation, the dataset is divided into *k* equal groups or folds. In this case five folds were used. Four groups are taken together and are used to train the data and one is used for validation. The procedure is repeated five times, each time leaving out a different set for validation as in Figure 4. The model error is estimated by averaging the errors of the five validation samples.

Figure 4: K-fold cross-validation sampling design.



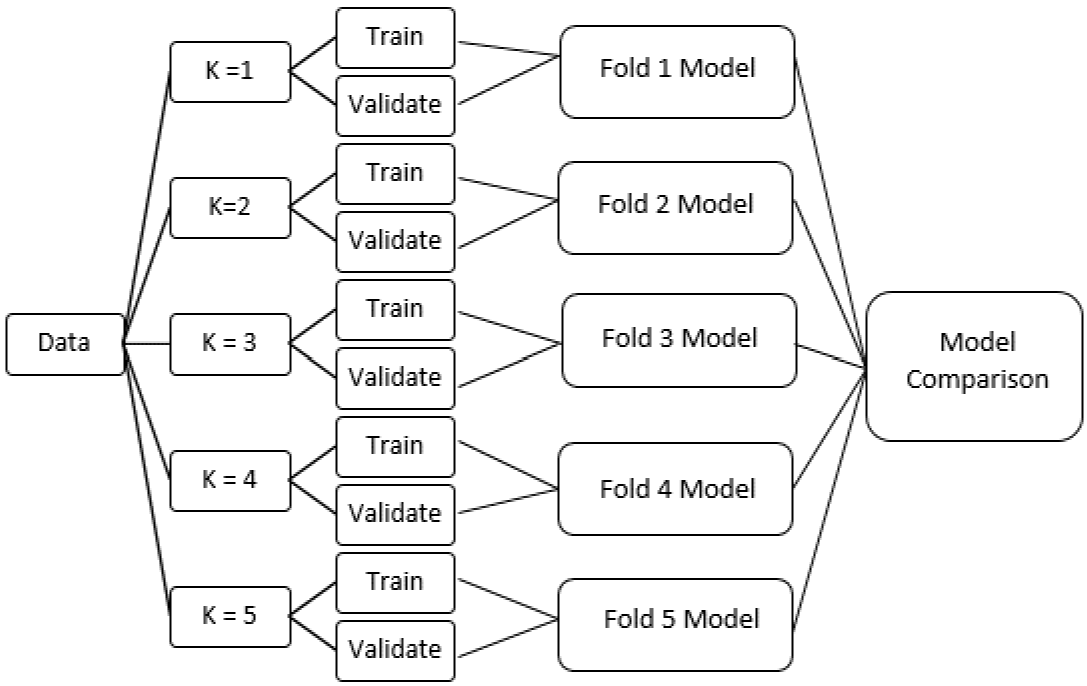
Five different modeling methods were tested and compared using k-fold cross validation. A general data mining diagram for running a modeling method with k-fold cross validation can be seen in Figure 5. Filters can be applied to select the proper groups for the validation and training sets for each fold, then the training and validation sets are sent to the modeling nodes where the same modeling method is run for each of the five training sets. The model is then run on each validation set for calculating the error. A model comparison node provides the relevant model evaluation statistics for each of the five folds.

The five different methods used to develop predictive models were: CHAID[2] (chi-square automatic interaction detection), BFOS-CART (the classification and regression tree method; Breiman, Friedman, Olshen, and Stone, 1984), a general decision tree, gradient boosting, and linear regression.   Each model was developed to predict the first semester GPA of the first-time full-time fall 2014 freshmen cohort. The average squared errors (ASE) of the five validation

[2] The CHAID and CART methods have been closely approximated by using Enterprise Miner settings. SAS Institute Inc. 2014. *SAS® Enterprise Miner™ 13.2: Reference Help*. Cary, NC: SAS Institute Inc. p. 755-758.

samples for each method were averaged and compared with the average errors of the training samples to check for overfitting and to find the method with the smallest error.

Figure 5. A general data-mining diagram for running 5-fold cross-validation to evaluate the accuracy of a model.



With the exception of linear regression, the methods tested were decision tree-based methods. The CART method begins by doing an exhaustive search for the best binary split. It then splits categorical predictors into a smaller number of groups or finds the optimal split in numerical measures. Each successive split is again split in two until no further splits are possible. The result is a tree of maximum possible size, which is then pruned back algorithmically. For interval targets the variance is used to assess the splits; for nominal targets the Gini impurity measure is used. Pruning starts with the split that has the smallest contribution to the model and missing data is assigned to the largest node of a split. This method creates a set of nested binary decision rules to predict an outcome.

Unlike CART with binary splits evaluated by the variance or misclassification measures, the CHAID algorithm uses the chi-square test (or the F test for interval targets) to determine significant splits and finds independent variables with the strongest association with the outcome. A Bonferroni correction to the p-value is applied prior to the split. CHAID may find multiple splits in continuous variables, and allows splitting of categorical data into more than two categories. This may result in very wide trees with numerous nodes at the first level. As with CART, CHAID allows different predictors for different sides of a split. The CHAID algorithm will halt when statistically significant splits are no longer found in the data.

The software was also configured to run a general decision tree that does not conform or approximate mainstream methods found in the literature. To control for the large number of nodes at each level, the model was restricted to up to four-way splits (4 branches), as opposed to CHAID which is finds and utilizes all significant splits and CART which splits each node in two. The F test was used to evaluate the variance of the nodes and the depth of the overall tree was restricted to 6 levels. Missing values were assigned to produce an optimal split with the ASE used to evaluate the subtrees. The software's cross validation option was selected in order to perform the cross validation procedure for each subtree. That results in a sequence of estimates using the cross validation method explained earlier to select the optimal subtree.

The final tree method was gradient boosting which uses a partitioning algorithm developed by Jerome Friedman. At each level of the tree the data is resampled a number of times without replacement. A random sample is drawn at each iteration from the training data set and the sample is used to update the model. The successive resampling results in a weighted average of the re-sampled data. The weights assigned at each iteration improve the accuracy of the predictions. The result is a series of decision trees, each one adjusted with new weights to improve the accuracy of the estimates or to correct the misclassifications present in the previous tree. Because the results at each stage are weighted and combined into a final model, there is no resulting tree diagram. However, the scoring code that is generated allows the model to be used to score new data for predicting outcomes.

The final method tested was linear regression. The discussion that follows highlights some of the difficulties in implementing linear regression in a data mining environment. Decision tree methods are able to handle missing values by combining them with another category or using surrogate rules to replace them. Linear regression, on the other hand, will listwise delete the missing values. Data in this study was obtained from multiple campus sources, and as such, many students did not have any records for some predictors. For example, students who did not apply for financial aid will have missing data on financial aid measures, a small percentage of the entering freshmen do not have SAT scores, and some students may not have courses utilizing the LMS. These measures result in an excessive amount of data being listwise deleted. The software has an imputation node that can be configured to impute missing data. For this study the distribution method was used whereby replacement values are calculated from random percentiles of the distributions of the predictors. There are many imputation methods and a thorough study of missingness for such a large number of variables is very time consuming. If the linear regression method appeared promising, other imputation methods would be explored and studied in greater detail. Another issue of concern in the linear regression analysis was multicollinearity. That is another issue that can take time to address thoroughly. For this analysis clustering was employed to reduce multicollinearity. With a large volume of predictors, it would be difficult and time consuming to evaluate all of the potential multicollinearity issues, so the software clustering node was used to group highly correlated variables. In each cluster, the variable with the highest correlation coefficient was retained and entered into the modeling process, and the others were eliminated.

Results

Gradient boosting had the smallest average ASE followed by that of CART (Table 1). Additionally, gradient boosting and BFOS-CART, on average, had the smallest differences between the validation and training errors. Those absolute errors were both approximately 0.02, while for the other methods it was greater than 0.1. Gradient boosting had the lowest average validation error, 0.375, while CHAID and linear regression had the highest at 0.49. Though gradient boosting had the lowest average validation ASE, the CART method was chosen for the modeling process. Close inspection of the CART results did not show evidence of any problems with the fit of the model, and it had a relatively low average ASE. The main reason for choosing the CART model is that gradient boosting, without an actual tree diagram, would make the results much more difficult to explain, use, and visualize. Having a set of student characteristics assigned to each node, as well as the ability to graphically display the decision tree adds to the

Table 1. Average Squared Error (ASE) Results for the Five Data Mining Methods

Data Mining Method	Traing and Validation ASE	K Folds					Average ASE
		1	2	3	4	5	
Gradient Boosting	Validation	0.333	0.353	0.377	0.391	0.422	**0.375**
	Training	0.363	0.358	0.351	0.351	0.343	**0.353**
BFOS-CART	Validation	0.394	0.425	0.429	0.436	0.525	**0.442**
	Training	0.427	0.423	0.432	0.433	0.393	**0.422**
CHAID	Validation	0.444	0.479	0.508	0.510	0.511	**0.490**
	Training	0.355	0.325	0.312	0.304	0.345	**0.328**
Decision Tree	Validation	0.421	0.432	0.472	0.495	0.515	**0.467**
	Training	0.335	0.330	0.325	0.304	0.312	**0.321**
Linear Regression	Validation	0.374	0.477	0.515	0.522	0.561	**0.490**
	Training	0.396	0.388	0.363	0.376	0.371	**0.379**

utility of the CART model. Once the CART method was selected, the model was run again using all of the data, and scoring output was created.

The score distribution table, Figure 2, which is part of the decision tree output allows us to view the frequencies of the model predictions. Twenty bins, the prediction ranges, are created by evenly dividing the interval between the lowest and highest predictions, 1.30 and 3.76. (Intervals without students are not listed.) The model score is calculated by taking the mid-point of the prediction range. The average GPA column contains the average GPA of the N students in the data that fall within the given range. The table can aid us in choosing GPA cut points for different interventions since it shows the number of students at the various prediction levels.

Table 2. Score Distribution Table

Prediction Range	Average GPA	N	Model Score
3.64 - 3.76	3.76	37	3.70
3.51 - 3.64	3.60	459	3.57
3.39 - 3.51	3.46	257	3.45
3.27 - 3.39	3.35	78	3.33
3.14 - 3.27	3.23	344	3.21
3.02 - 3.14	3.08	665	3.08
2.90 - 3.02	2.93	478	2.96
2.65 - 2.78	2.74	89	2.71
2.53 - 2.65	2.61	362	2.59
2.41 - 2.53	2.52	16	2.47
2.04 - 2.16	2.12	18	2.10
1.92 - 2.04	1.94	25	1.98
1.55 - 1.67	1.59	13	1.61
1.30 - 1.43	1.30	11	1.36

Table 3. Variable Importance Table.

Variable	Relative Importance
High School GPA	1.0000
Scholarship Aid (Yes/No)	0.9643
Total AP non-STEM course accepted for credit	0.8980
Total AP STEM course accepted for credit	0.8729
LMS logins per STEM course weeks 2 -6	0.8619
Total LMS STEM course logins, weeks 2 -6	0.8542
LMS logins per non-STEM course, weeks 2 -6	0.8214
Area of residence at time of admission	0.7921
Total LMS non-STEM logins, weeks 2 – 6	0.7888
Student has a declared major or area of interest	0.6902
Total fall 2014 non-STEM enrolled units	0.6859
Total LMS non-STEM course logins, week 1	0.6712
Total fall 2014 STEM enrolled units	0.5789
Avg. SAT Math-CR-Writing score of the high school	0.5577
Student SAT Math-CR	0.5540
Avg. SAT CR score of the high school	0.5357
Total LMS STEM course logins, week 1	0.5307
Avg. SAT Math-CR score of the high school	0.5176
Total STEM courses	0.5119
Avg. SAT Math score of the high school	0.5080
Total non-STEM courses	0.4808
Type of math course in term 1 (e.g., pre-college, calculus level)	0.4636
Total STEM courses using LMS	0.4258
Advising visits, week 1 pertaining to registration	0.3826
Ethnic group	0.3609
Highest DFW rate in non-STEM course	0.3425
Student SAT Math score	0.3197
Total non-STEM courses using LMS	0.3115
Total Athletics Aid	0.2736
Total high DFW STEM enrolled units	0.2714
Intramural sports participation	0.2548
Tutoring Center visits for STEM courses, weeks 1 – 6	0.2533
Fitness Class attendance	0.2378
Student SAT CR score	0.2146
Highest DFW rate for enrolled STEM course	0.1868
Honors College or Women in Science & Eng. (Yes/No)	0.1827
Total high DFW enrolled STEM courses	0.1624
Stony Brook Math Placement Exam score	0.1500
Student SAT Writing Score	0.1495
Total grant aid	0.1436
% of freshmen in student's highest DFW rate STEM course	0.1191
Total loans distributed (per Fin. Aid Off. Records)	0.1155
Advising visit during week 1, not registration-related	0.1149
% of 1st years in student's highest DFW rate non-STEM course	0.0721

Table 3 lists the relative importance measure for variables that were entered into the modeling process. The relative importance measure is evaluated by using the reduction in the sum of squares that results when a node is split, summing over all of the nodes.[3] In the variable importance calculation when variables are highly correlated they will both receive credit for the sum of squares reduction, hence the relative importance of highly correlated variables will be about the same. For that reason some measures may rank high on the variable importance list, but do not appear as a predictors in the decision tree.

On Table 3 high school GPA is highest on the variable importance list for predicting freshmen GPA when modeled mid-semester, followed by whether or not a student received a scholarship. Next are AP STEM and non-STEM courses accepted for credit, and then LMS system logins. A student's combined SAT Math and Critical Reading Exam Score is 15th on the list just behind the high school average score for the combined SAT Math, Critical Reading, and Writing exam. Some other measures that exceeded SAT scores in relative importance are whether a student has a declared major, and the geographic area of residence when admitted.

The decision tree generated by the model is presented in two parts in Figures 6 and 7. The CART method, employing only binary splits as previously discussed, selected high school GPA for the first branch of the tree modeling first semester freshmen GPA. High school GPA was split into two nodes, less than or equal to 92.0, and greater than 92.0 or missing. Figure 6 displays the portion of the decision tree with high school GPA less than or equal to 92.0 and Figure 7 has the portion of the tree with high school GPA greater than 92.0 or missing.

Figure 6. Part 1 of the CART Decision Tree Model Predicting Freshmen GPA for Students Having a High School GPA <= 92.0.

[3] . SAS Institute Inc. 2014. *SAS® Enterprise Miner™ 13.2: Reference Help*. Cary, NC: SAS Institute Inc. p. 794.

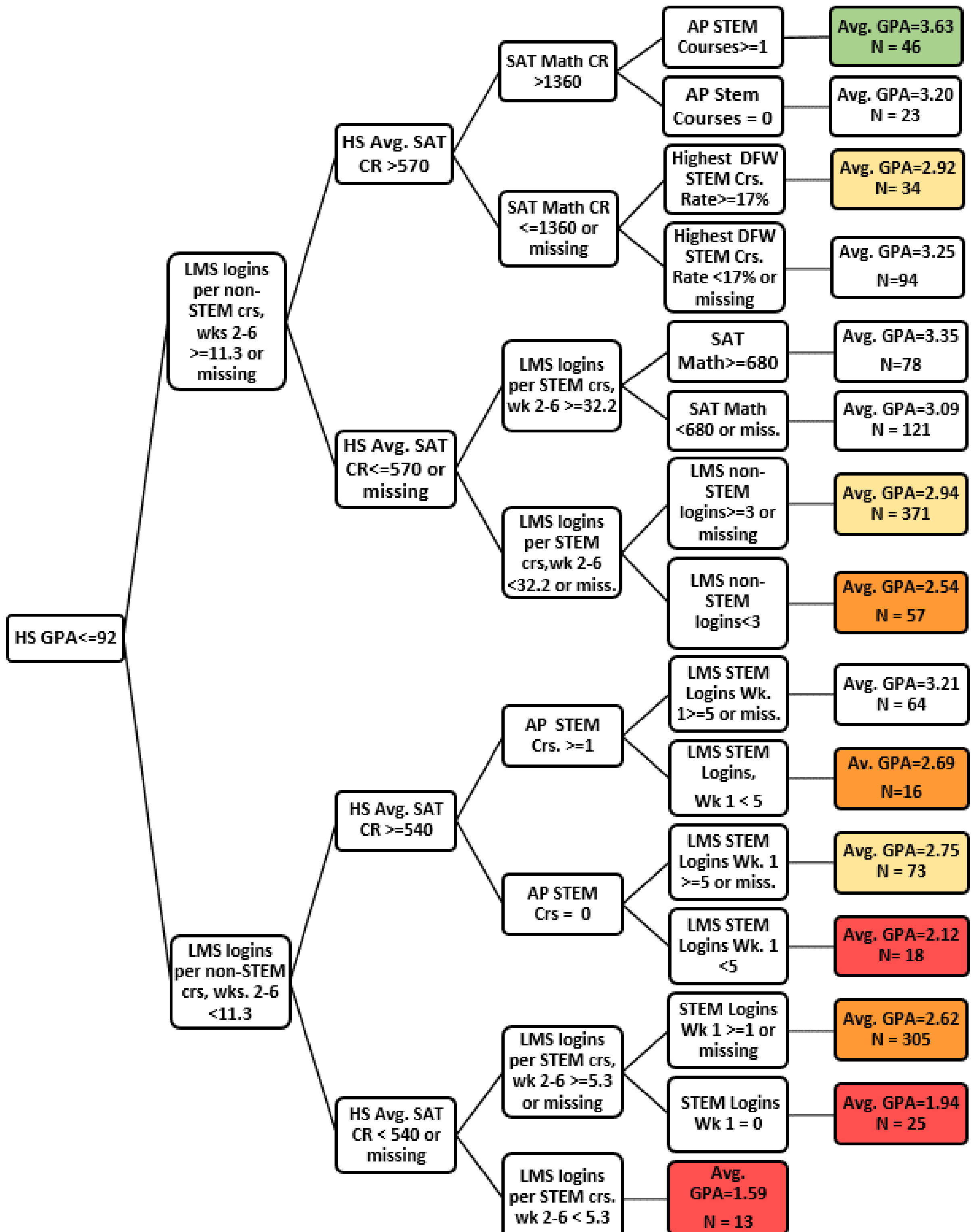
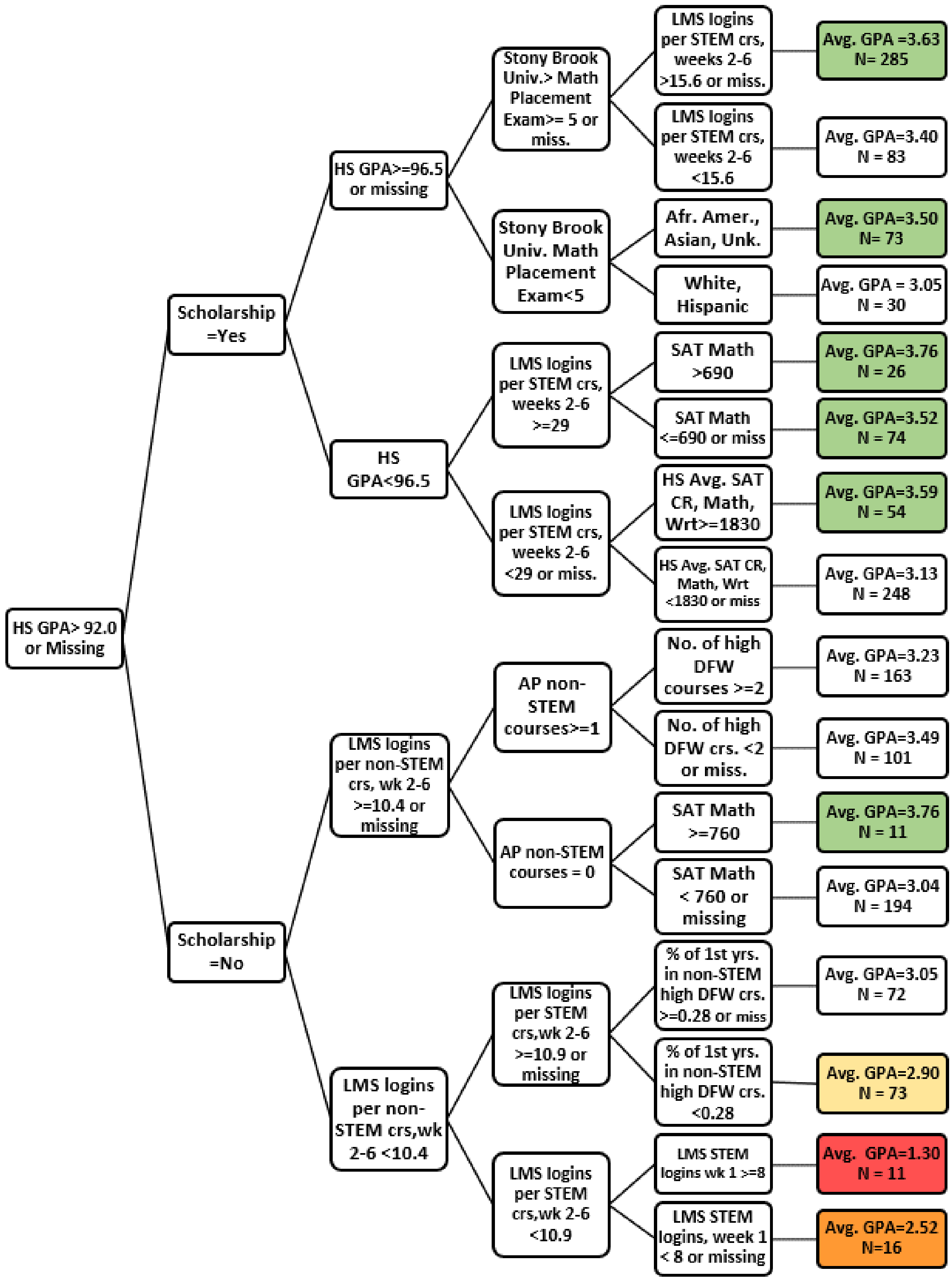


Figure 7. Part 2 of the CART Decision Tree Model Predicting Freshmen GPA for Students Having a HS GPA > 92.0 or missing.



The next branch for the lower high school GPA group is the non-STEM course LMS logins during weeks 2 through 6. Average high school SAT scores appear at the next level.

Figure 7 displays the section of the tree having the students with a high school GPA greater than 92.0 or missing. A small number of students, some of them international students, do not have a high school GPA in their records. The CART algorithm has combined those observations with the node having high school GPA > 92.0. In that way, those observations remain in the model and are not listwise deleted as they would be in a standard linear regression analysis. The next two levels are different than those for the lower high school GPA students. The next split after high school GPA is whether the students received a scholarship or not. For those who received a scholarship another high school GPA node follows that splits the students into groups above and below 96.5, while for those without a scholarship LMS non-STEM logins during weeks 2 through 6 is most important

Examining both sections of the tree in Figures 6 and 7, we see that LMS logins factored in numerous splits confirming that students' interactions with the college environment plays a role in their academic success. We also observe the differences in the decision rules for students in the higher high school GPA group as compared to the students in the lower high school GPA group.

The actual GPA predictions can be found in the nodes in the right-most column of the tree and are the average GPA's of the students represented by the characteristics of each particular node. The characteristics associated with the GPA predictions can be ascertained by tracing the paths from the high school GPA node on the left to the desired average GPA node on the right. For example, to determine the characteristics for the students represented in the top right average GPA = 3.63 node in figure 6, we have students with high school GPA < =92, LMS logins per non-STEM course in weeks 2 to 6 >= 11.3 or missing, high school average SAT critical reading > 570, SAT Math – Critical Reading combined score > 1360, and finally, receiving credit for 1 or more AP STEM courses. The prediction, 3.63, is the actual average GPA of students in the fall 2014 cohort having the characteristics just listed. Hence, we can say that students with characteristics represented in the final nodes have, *on average*, the GPA that is listed in the node.

The average GPA nodes have been color-coded to assign estimated risk to the GPA levels. The red nodes have average GPA's of 2.20 or less and are at the *highest risk* of receiving a low GPA The orange nodes represent *high risk* students and on average have GPA's of above 2.20 to 2.75. Yellow nodes with average GPA's of above 2.75 to 3.0 represent *moderate risk,* white nodes represent neutral risk with average GPA's ranging from above 3.0 to below 3.5, and the green nodes are *low risk* students who, on average, have GPA's of 3.5 and above. The given risk levels can be adjusted based on university outcomes and the number of students assigned to various planned interventions.

Conclusion

It is clear from studying the decision tree model that weaker students from high schools with lower average SAT scores, who additionally are interacting with the LMS at diminished rates are over-represented in the lower GPA groups. The model can assist in identifying these students before the end of the semester so they can be assigned to interventions that may help to improve their outcomes. Since enrollment in courses with higher failure rates is also a factor appearing in the decision tree, developing a pre-orientation model could assist advisors in steering some students from course loads that may be excessively burdensome. The model results can also be shared with departments to inform their advising and intervention efforts. Automated methods for easily sharing the results are being planned. The goal is to find the students who need assistance in fulfilling their potential, thereby reducing the number who end up leaving due to poor performance.

References

Bahr, P.R. (2008). Does mathematics remediation work?: a comparative analysis of academic attainment among community college students. *Research in Higher Education*. 49:420-450.

Bean J. (1983). The application of model of turnover in work organizations to the student attrition process. *Review of Higher Education*, 6(2), 129-148.

Breiman, L., Friedman, J., Olshen, R., Stone, D. (1984): *Classification and Regression Trees*. Wadsworth Books.

Chen R. (2012). Institutional characteristics and college student dropout risks: a multilevel event history analysis. *Research in Higher Education*. 53:487–505.

Friedman, J., Hastie, T., Rosset, S., Tibshirani, R., Zhu, J. (2003) Discussion of Boosting Papers. Retrieved from http://web.stanford.edu/~hastie/Papers/boost_discussion.pdf

Herzog S. (2005). Measuring determinants of student return vs. dropout/stopout vs. transfer: a first-to-second year analysis of new freshmen. *Research in Higher Education*. 46:883-928.

Johnson I. (2006). Analysis of stop-out behavior at a public research university: the multi-spell discrete-time approach. *Research in Higher Education*. 47:905-93.

Johnson I. (2008). Enrollment, persistence and graduation of in-state students at a public research university: does high school matter? *Research in Higher Education*. 49:776-793.

Parker M. (2005). Placement, retention, and success: a longitudinal study of mathematics and retention. *The Journal of General Education*. 54:22-40.

Robbins S, Allen J, Casillas A, Akamigbo A, Saltonstall M, Campbell R, Mahoney E, Gore P. (2009). Associations of resource and service utilization, risk level, and college outcomes. *Research in Higher Education*. 50: 101-118.

SAS Institute Inc. (2014). *SAS® Enterprise Miner™ 13.2: Reference Help*. Cary, NC: SAS Institute Inc.

Singell L, Waddell, GR. (2010). Modeling retention at a large public university: can at-risk students be identified early enough to treat? *Research in Higher Education*. 51:546-572.

Stater M. (2009). The impact of financial aid on college GPA at three flagship public institutions. *American Educational Research Journal.* 46:782-815.

Stinebrickner R, Stinebrickner T. (2014). Academic performance and college dropout: using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*. 32:601-644.

Thomas EH, Galambos N. (2004). What satisfies students? mining student-opinion data with regression and decision-tree analysis. *Research in Higher Education*. 45:251-269.

Tinto, V. (1987). Leaving college: rethinking the causes and cures of student attrition. Chicago, IL: The University of Chicago Press.

Zwick R, Sklar JG. (2005). Predicting college grades and degree completion using high school grades and SAT scores: the role of student ethnicity and first language. *American Educational Research Journal*. 42:439-464.

Appendix

Variable List

<u>Demographics</u>
Gender
Ethnicity
Area of residence at time of admission: Suffolk County, Nassau County, New York City, other NYS, other US, International

<u>Pre-college Characteristics</u>
High School GPA
College Board SAT Averages by High School
- Average High School Critical Reading
- Average High School SAT Math
- Average High School SAT Critical Reading + Math

SAT: Math, Critical Reading, Writing, Math+Critical Reading

<u>College Characteristics</u>
Number of AP STEM courses accepted for credit
Number of AP non-STEM courses accepted for credit
Total credits accepted at time of admission
Total STEM courses
Total STEM units
Total Non-STEM courses
Total No-STEM units
Class level
Dorm Resident
Intermural Sports Participation
Fitness Class Participation
Honors College
Women in Science and Engineering
Educational Opportunity Program
Stony Brook University Math and Writing Placement Exams
College of student's major or area of interest: Arts and Sciences, Engineering, Health Sciences, Marine Science, Journalism, Business
Major Group: business, biological sciences health sciences, humanities and fine arts, physical sciences and math, social behavioral science, engineering and applied sciences, journalism, marine science, undeclared, other
Major type: declared major, undeclared major, area of interest
High DFW Rate Courses: enrollment >= 70, percent DFW >=10%
- Total high DFW STEM units
- Total high DFW non-STEM units
- Highest DFW rate among the DFW Courses in which the student is enrolled
- Highest DFW rate among the DFW Courses in which the student is enrolled

Proportion of freshmen in a student's highest DFW rate STEM course
Proportion of freshmen in a student's highest DFW rate non-STEM course
Type of math course: high school level, beginning calculus, sophomore or higher math

Financial Aid Measures
Aid disbursed in the Fall 2014 – Spring 2015 academic year
Total grant funds received
Total Loans recorded by the Financial Aid Office
Total scholarship funds received
Total work study funds received
Total athletics aid received
Athletic aid, grant, loan, PLIS loan, subsidized/unsubsidized loan, scholarship, work study, TAP, Perkins, Pell indicators
Adjusted Gross Income
Federal Need
Federal Expected Family Contribution
Dependent status

Services/Learning Management System (LMS)
Advising Visits/Tutoring Center Usage
Tutoring center appointment no shows
Number of STEM Course Center Visits, weeks 1 to 6
Number of non-STEM Course tutoring Center visits, weeks 1 to 6
Advising Visits during week 1
Advising visits during weeks 2 – 6
Course Management System Logins
F14_Stem_Login_N
F14_NonStem_Login_Week1_N
Non-STEM course related logins during weeks 2 - 6
Non-STEM Course related logins during week 1
STEM Course related logins during week 1
STEM Course related logins during weeks 2 to 6
Number of STEM course logins per STEM course using the CMS, weeks 2 – 6.
Number of non-STEM course logins per non-STEM courses using the CMS, weeks 2 – 6.

D R A F T

TITLE

Institutional Ethnography:
A Methodology for the Study of Inequality

AUTHOR

Elizabeth Seton Mignacca, PhD

TITLE

Institutional Ethnography:
A Methodology for the Study of Inequality

ABSTRACT

Institutional Ethnography (IE) is a method of inquiry used by researchers in the social sciences, human services, and in policy research. Validating experience as a form of knowledge, IE researchers are concerned with issues of social justice, the critique of objectified knowledge, and "mapping" how social relations can cause oppression in society. In doing so, the findings of IE studies point to possible interventions and transformations. This paper argues that IE methodology (Smith 2005, 2006) is especially relevant for institutional researchers working in Colleges and Universities with social justice missions, such as community colleges and Catholic universities.

KEY WORDS: Institutional Ethnography, Institutional Research, Inequality

Can institutional research conducted via feminist methods create more socially just institutions? This paper examines the contributions of feminist methods to research and considers the application of institutional ethnography (Smith 2005, 2006) to institutional research. I argue that institutional ethnography (IE), a mode of inquiry aimed at "mapping" how inequality is maintained through the social organization of institutions, is particularly well-suited for institutional research within colleges and universities with social justice missions, such as community colleges and Catholic universities.

Contributions of Feminist Methods

Feminist methods for social research have challenged assumptions of universal concepts and essential categories, contending these consist of ideas, practices and policies largely formed by dominant and privileged groups. Using poststructural concepts to critically examine the formal and informal policies and practices embedded and codified in research communities, they have enlarged the understanding of social inequality and social injustice. They have also made substantive contributions to methodology. Six main characteristics identified by Sielbeck-

Bowen et al. (2002) summarize the contributions of the feminist paradigm to social inquiry.

They include:

- focusing on inequalities that lead to social injustice
- enlarging the descriptive and analytic understanding of the systematic and structural nature of discrimination based on gender, race and class
- articulating the political nature of social inquiry and advocating for transparency in acknowledging the political commitments of researchers
- recognizing that knowledge is a resource that should be shared with those under study
- broadening the understanding of values as culturally, socially and temporally constructed
- acknowledging there are multiple ways of knowing, and that some ways are privileged over others (Sielbeck-Bowen et al. 2002)

By writing about the systematic and structural nature of gender inequalities, the feminist paradigm in research provides a model to broaden the discussion of racial, ethnic and class inequalities and the role institutions have played in maintaining them. It was out of these distinct characteristic that Institutional Ethnography developed.

Institutional Ethnography

In 1995 the American Sociological Association (ASA) honored the sociologist Dorothy Smith for the development of a mode of inquiry aligned with feminist principles of research which she called “institutional ethnography (IE).” Since that time, IE has come to be known for its democratic ethic and is now called, “a sociology for people” (Smith 2005). The methodology has been used by researchers working in human services, the social sciences and in policy research. One of the innovations of IE is that it sets aside theory at its outset and instead begins with descriptions of people’s everyday lives. In doing so, it shifts the focus of research from objective knowledge and theories of social problems onto how inequalities and institutional contradictions impact the lives of people. For institutional researchers in higher education, who may collect data and then attempt to “cut” it by race, class or gender, this marks a departure from traditional methods. In this way, IE provides a strategy for investigating differences traditional theory may have missed.

Data collection centers on methods of in-depth interviewing, observation and textual analysis -- those largely consistent with qualitative methods and scholars have noted IE commonalities with global ethnography, multi-site ethnography and political ethnography, which also center on inequalities (Bisaillion and Rankin 2013). However, IE’s distinction is that unlike anthropological ethnography, it is committed to revealing how official texts--broadly defined--come to shape the social relations of institutions. Using institutional texts, documents, forms and definitions, researchers can analyze how people gain access, participate in, and work within institutions. IE seeks to investigate how official ideologies embedded in these texts impact the social relations of an institution.

Researchers who have adopted IE principles explicitly seek to produce and distribute knowledge more democratically, so as to challenge inequality and highlight how things might be changed. In this sense, scholars have noted the humanistic nature of IE findings and how they can be employed by civil society and social justice advocates to help change policy or administrative practice (Society for the Study of Social Problems 2015).

Some of the main characteristics of IE include:

1. The development of a study “problematic” from the experiences of people’s everyday lives, instead of theorized or official definitions of problems.
2. Shifting from those experiences to how circumstances are “socially organized” by “mapping” what actually happens in the process of people gaining access to, participating in, and working within institutions.
3. Analysis identifies how texts mediate power through institutional forms of knowing and its impact on people’s lives.

4. A commitment to the principle that study findings should not only contribute to theory-building and research communities, but to educating the population of the institutions under study and those they serve. (Campbell and Gregor2004).

Like all ethnographies, IE studies typically produce descriptive findings of how people gain access to and participate in institutions. They also provide detailed understandings of how administrative practice is carried out, including what assumptions institutionalized work makes about the population being served. Some IE studies will produce diagram "maps" that display the movement of regulatory, legal, or dominant cultural ideology through administrative practice and the processing of paperwork.

This leads to the main question of this inquiry: Are IE methods particularly suited to study institutions that address social justice in their mission?

Mapping Inequality in Higher Education Institutions

Community Colleges are unique in the higher education sector for their focus on the local communities they serve. Historically, their mission has been to provide regional communities with geographic, academic and economic access to higher education (Beebe 2015). Despite these aims, the institutional intersection of legislation, policy, administrative practice and the "life chances" of community college students can work against these goals. Critics have sited the low completion and transfer rates of students (Rosenbaum, Deil-Amen, Person 2006) while others have studied an outdated and difficult to implement funding model (Goldrick-Rab 2010).

Catholic universities distinctly recognize the dignity of each person and strive to provide education that grounds students in the ethos and values of service, and which aligns to the belief of the oneness of the human family (Estanek, James and Norton 2006). Goals which often fly in the face of popular rhetoric that places the value of a college degree within the matrix of labor force projections.

In these contexts, institutional researchers might take the lead in bringing the experiences of marginalized groups to bear on how higher education formulates problems and organizes administrative work. Using IE principles, institutional researcher in these settings might begin to "map" a bigger picture of what higher education expects of non-dominant groups to be successful in our institutions. We might consider that these student populations face particular social injustices that are beyond the scope of the unitary "student" experience, so prevalent in our institutional research data.

Producing and Distributing Institutional Research More Democratically

As this paper is being written, community colleges are grappling with issues of social class representation in their governing boards (Smith 2015) and escalating tensions between minority student activists and university administrators have resulted in a system president and flagship university chancellor's resignation (Woodhouse 2015). What modes of inquiry might help institutions better understand these tensions? What is the role of institutional researchers in the development of positive solutions? How should institutions with social justice missions contribute to formulating data that would support positive change? This paper does not provide the answers to these dilemmas, but urges institutional researchers to raise questions and to consider how feminist research principles may assist institutions in realizing their social justice missions.

REFERENCES

Beebe, Anthony. 2015. “Serving Social Justice.” *Community College Journal*. 86(2):48-53.

Bisaillon, Laura and Janet Rankin. 2013. “Navigating the Politics of Fieldwork Using Institutional Ethnography: Strategies for Practice.” *Forum: Qualitative Social Research*, 14(1).

Campbell, Marie and Frances Gregor. 2004. *Mapping Social Relations: A Primer in Doing Institutional Ethnography.* Lanham, MD: AltaMira Press.

Estanek, Sandra M., Michael J. James, and Daniel A. Norton. 2006. “Assessing Catholic Identity: A Study of Mission Statements of Catholic Colleges and Universities.” *Catholic Education: A Journal of Inquiry and Practice,* 10(2) 199-217.

Goldrick-Rab, Sarah. 2010. “Challenges and Opportunities for Improving Community College Student Success. *Review of Educational Research*. 80(3): 437-469.

Rosenbaum, James E., Deil-Amen, Regina, and Ann E Person. 2006. *After Admission: From College Access to College Success*. New York, NY: Russell Sage.

Sielbeck-Bowen, Kathryn A. and Sharon Brisolara, Denise Seigart, Camille Tischler and Elizabeth Whitmore. 2002. “Exploring Feminist Evaluation: The Ground from Which We Rise.” *Special Issue: Feminist Methods for Evaluation. New Directions for Evaluation*, 96: 3-8.

Smith, Ashley A. 2015. “Board Tussle Over Race.” Inside Higher Ed. Downloaded November 5, 2015: https://www.insidehighered.com/news/2015/11/05/milwaukee-technical-college-board-under-fire-lack-minority-representation

Smith, Dorothy. 2005. *Institutional Ethnography: A Sociology for People* (Gender Lens Series) Lanham, MD: AltaMira Press.

Smith, Dorothy (Ed). 2006. *Institutional Ethnography as Practice*. Lanham, MD: Rowman & Littlefield Publishers.

DeVault, Marjorie (Ed). 2015. Society for the Study of Social Problems Institutional Ethnography Division Mission Statement. Downloaded November 10, 2015: http://www.sssp1.org/index.cfm/pageid/1236/m/464

Woodhouse, Kellie. 2015. “Activist Oust Two Leaders.” *Inside Higher Ed*. Downloaded November 10, 2015: https://www.insidehighered.com/news/2015/11/10/u-missouri-leaders-resign-amid-student-concerns-over-racism-and-diversity

UNDERSTANDING THE IMPACTS OF THE TEST OPTIONAL ADMISSION POLICY

Yuko Mulugetta, Ph.D.

Director of Enrollment Planning
Division of Enrollment Management
Ithaca College

Over 850 institutions have adopted a "test-optional policy" (TOP), promoting campus diversity by removing the barriers against minority groups often presented by standardized testing. The TOP and diversity causal relationship, however, is not well researched. Using six cohorts' data from Ithaca College, this study employs a quasi-experimental design and reveals that the treatment group (non-test submitters) did indeed increase the probability of a student being a minority, controlling for non-TOP changes observed in the two control groups (test submitters) before and after TOP implementation. TOP positively affected diversity at each stage of the enrollment funnel: application, admission, enrollment and retention.

Introduction

The test optional movement continues to gain popularity among enrollment officers. By 2015, over 850 colleges and universities, including several well-known national universities such as Wake Forest and George Washington (FairTest, 2015), have adopted a "test-optional policy" (TOP). This policy enables students to opt out of submitting standardized test scores as a part of their admission applications. Ithaca College, a mid-sized four-year residential comprehensive college in central New York, is one of those TOP institutions. The College implemented the policy in 2012 for the admission applications of the fall 2013 entering cohort.

One of the main goals underlying Ithaca's decision was to increase campus diversity by removing the barriers against minorities often presented by standardized testing. Here, a minority group member is broadly defined as a member of a racial minority, a certain socio-economic group (i.e. Pell recipient), a first-generation college student, or a student with learning differences (LD). While many administrators of TOP have presented anecdotal information asserting that the policy has increased campus diversity (Simon, 2015; Rochon, 2013), in-depth research on the impact of the test optional admission policy is still in its early stages. It is, therefore, of paramount importance that the institutional researchers of those TOP institutions that have already implemented the policy should conduct research on its impact, provide data-informed support for future decisions, and share their findings with other institutions. The present study is an effort to provide other institutions with a research example so that more research results can be compiled and shared to advance our understanding of the impact of the test optional admission policy.

Literature Review

The controversy over the validity of the use of standardized test scores in the college admission process is nothing new. The early intent of the creation of the SAT was to open the doors of higher education to students without the traditionally-valued credentials; the standardized testing scheme was seen as a way to "level the field". Along with this motivation, colleges and universities also saw standardized testing as a way to enhance their prestige by showing that their students were highly qualified based on test results -- not based on social class or connections (Epstein, 2009). The premise that standardized testing can effectively identify qualified students and accurately predict their future academic success justified use of these tests and led to their dominating the college admissions world in the latter half of the 20th century.

This premise, however, has become subject to severe scrutiny in recent years. The main criticism is that standardized tests are culturally biased against subgroups including racial minority groups, females, first generation students, and those from low-income strata (e.g., Zwick, 2004, 2007). Empirical studies have revealed that female students' SAT math scores are lower than males by one-third of a standard deviation while Latinos' and Afro Americans' scores are lower than whites by two-thirds and one standard deviation respectively (Rosner, 2012).

The critics argue, therefore, that standardized tests structurally maintain -- or worse augment -- the already existing gap between advantaged and disadvantaged applicants, by imposing “a devastating impact on the self-esteem and aspirations of young students” (Atkinson, 2001).

Furthermore, it has been argued that standardized test measures are not only culturally biased, but that they are also not the best predictor of future academic achievement in college. The studies have consistently found that SAT scores do not predict the college first-year GPA as effectively as other measures such as high school GPA or AP credits (e.g., Cornwell, Mustard, and Van Parys, 2012; Wonnell, Rothstein, and Latting, 2012; Rask and Tiefenthaler, 2012). The College Board research team has examined the incremental validity attributed to SAT scores over high school GPA (HSGPA) in predicting the first-year college GPA (FYGPA). The study used a large cross-sectional sample of data from the 2006 cohort who took the revised SAT with the newly added SAT writing section. They found that when HSGPA was taken into account, the incremental validity attributed to SAT scores was 0.08, which was lower than the incremental validity associated with HSGPA over SAT scores ($r = 0.09$). Because of these results, they recommended that colleges use both HSGPA and SAT scores to make the best predictions of student success (Kobrin, Patterson, Shaw, Mattern, and Barbuti, 2008).

The recent research conducted by Ithaca College (Mulugetta, 2013) using the hierarchical regression technique, has clearly shown that standardized tests add surprisingly small explanatory power after HSGPA and AP credits are considered in predicting students’ academic performance in college. Conversely, strength of schedule along with HSGPA and AP credits were found to be critically important in the admission process. Based on this research, Ithaca College implemented the TOP policy in 2012 for admission applications of the fall 2013 entering cohort. This Ithaca College study stated that “the non-SAT measures seem to play a particularly significant role in admitting qualified students from minority groups." This follow-up study is an effort to conduct an in-depth investigation of the impacts of the TOP on campus diversity using Ithaca's six cohorts' data.

Test Optional Policy Controversy

While in-depth research on the impact of the test optional admission policy is still in its nascent stages, two landmark studies were published in 2014, which have sparked heated national debate on the TOP impact on educational outcomes and campus diversity.

"Defining Promise: Optional Standardized Testing Policies in American College and University Admissions" by William C. Hiss and Valerie W. Franks examined 122,916 student and alumni records of eight cohorts (2003 to 2010) provided by a wide variety of four-year institutions, including twenty private institutions, six public universities, five minority serving institutions and two arts schools, which represented twenty-two U.S. states and territories. Analyzing this wide-range of national data, the authors focused on one simple, but fundamental question: "Are college admissions decisions reliable for students who are admitted without SAT or ACT scores?" The study answers the question affirmatively revealing that the academic outcome difference between test-submitters and non-submitters was .05 of cumulative GPA (2.88 vs. 2.83 respectively) and 0.6 percent in graduation rates (63.9% vs 63.3% respectively), concluding "By any standard, these are trivial differences." This study has confirmed the findings of other studies that high school GPA is the best predictor of college GPA.

Hiss and Franks found that non-submitters are more likely to be first-generation college students, all categories of racial minorities, women, Pell Grant recipients, and students with learning differences (LD). Furthermore, the study pointed out that white students also opted out of test score submissions at rates within low double digits of the average. Interestingly, the study discovered a bimodal income distribution among non-submitters; on one hand, the financially needy group that consisted of first-generation, minority students and Pell Grant recipients, and on the other, the no need group who did not request financial aid. The authors mentioned that non-test submitters are often not considered for awards based solely on merit despite their high achievements because many institutions require test scores for merit awards consideration.

In conclusion, "Defining Promise" asks one last question: "Does standardized testing produce reliable predictive results, or does it artificially truncate the pool of applicants who would succeed if they could be encouraged to apply?" The authors have firmly stated, "At least based on this study, it is far more the latter."

Hiss and Frank's three-year national study has opened an exciting new chapter for test optional research and proposed a number of important topics for future studies. One of them is how the TOP impacts each stage of the admissions funnel. The authors wrote:

> "While the number of private institutions with optional test policies continues to expand modestly, the share of [enrolled] students within these institutions choosing to be non-submitters is also climbing over time … We did not gather data to analyze admissions funnels, so do not know whether this increased share of non-submitters is due to larger pools of non-submitter applicants from which to choose stable enrollments, higher yield from admissions offers to non-submitters, reshaped admissions priorities by admissions staffs, or colleges using non-submitter applications to increase overall enrollments. As with several other facets of this study, the admissions funnel data is a promising topic for further study (p. 12)."

The other landmark study, "The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boom for Equity or Something Else?" by A. S. Belasco, K. O. Rosinger and J. C. Hearn (2014), separately investigated how the TOP affected application and enrollment at different points of the admissions funnel, using the institutional level panel data of 180 selective liberal arts colleges including 32 TOP institutions from 1992 to 2010. The core question of their research was whether the TOP adoption did in fact increase low-income and racial minority student enrollment, or whether the TOP institutions simply accomplished the goal of raising their institutional status in the form of greater application numbers and higher reported test scores. The study carefully isolated plausible causal factors by employing a quasi–experimental design with the treatment group (TOP institutions) and the control group (non-TOPs), and applying the DiD (Difference in Difference) statistical analysis technique. The study found that the TOP institutions failed to demonstrate a positive change in the proportion of low-income and minority student enrollment after controlling institution-specific and year-specific effects. On the other hand, it shows that the TOP did indeed benefit the institutions by increasing the number of applications thus becoming more selective, and by raising their reported SAT scores significantly (about 26 points). The authors wrote:

> "Despite the clear relationship between privilege and standardized test performance, the adoption of test-optional admissions policies does not seem an adequate solution to providing educational opportunity for low-income and minority students. In fact, test-optional admission policies may perpetuate stratification within the postsecondary sector, in particular, by assigning greater importance to credentials that are more accessible to advantaged populations." (p. 13)

Obviously, two studies, using different research approaches and data, have reached contradictory findings: Hiss and Frank have discovered TOP's positive role in encouraging diverse groups of students to enroll and succeed at college, while Belasco and others did not find the evidence to reveal an affirmative impact of the TOP on enrollment diversity.

The TOP Impacts on the Enrollment Funnel

The main purpose of this study is to provide insights into how the TOP affects the diversity of the student body at <u>each point</u> of the enrollment funnel using Ithaca College's data as an example. As Hiss and Frank have correctly pointed out, it is critical to know how the TOP is affecting the student profile and composition at each stage of the enrollment funnel and to understand what other interactive factors are driving that phenomenon.

While the enrollment community has long studied and debated the definition and the importance of "the funnel", the present study simply views the enrollment funnel as "a foundational mechanism to represent the prospective student pipeline" (Copeland, 2009) through which a prospective student makes a series of complex decisions as s/he progresses down the path to enrollment and ultimately graduation. The present study intentionally calls the funnel "the enrollment funnel" instead of "the admission funnel" since this author believes the ultimate goal of the funnel as not merely enrolling capable students, but graduating them from the institution.

Many enrollment professionals view that this pipeline is composed of various stages each of which is characterized by its own decision-driven actions. The prospective students are labeled at each stage as: Suspects who are potential students; Inquirers who have expressed interest in admission; Applicants who have submitted applications for admission; Admits who are accepted for admission; Paids who have submitted enrollment deposits; and Enrolled who have actually registered and attended courses at the institution. This study adds two more levels to the funnel: Retained who have persisted at the institution, and Graduated who have completed the requirements and obtained a degree from the institution. We must acknowledge that a student's actions going through the pipeline involve very complex decision-making processes influenced by many factors. Examples of these factors are: educational quality and reputation of the institution, academic programs available, financial aid offers, perceived value of its education compared to competing institutions, influence/advice of social networks (parents, peers, guidance counselors, athletic coaches etc.), environmental factors such as weather or location, and distance from home.

As Hiss and Frank have stated, it is of paramount importance to know how the TOP is affecting both the students' and the institution's decision-making at each stage of the enrollment funnel. Independent of the Hiss study, Belasco and others from University of Georgia attempted to answer this funnel question, but the their study looked at the TOP influence only at the application and enrollment stages, and ignored the most critical stage: admission. The study failed to investigate how the TOP affected the diversity of the applicants who were accepted by the institution. The present study attempts to show that addressing this critical funnel stage will expand and deepen our understanding of the impact of TOP on the campus landscape. For example, it would be useful to know if the TOP can positively affect diversity among students who have applied and been admitted, but not positively affect the diversity among enrolled and/or retained students. If this is the case, we should ask ourselves what other interactive factors may be preventing the accepted TOP students from enrolling. One factor could be diminished opportunity for merit awards for non-test submitters who are often excluded from the merit award selection process as Hiss and Franks stated in their study.

The present study is a first attempt to provide insights into how the TOP affects the diversity of the student body at each of the four stages of the enrollment funnel: application, admission, enrollment, and retention.

Research Goals

The present study analyzes 90,824 individual applicant records from the three test-optional cohorts and the three cohorts prior to the implementation of TOP. The study defines a minority group member as a member of a racial minority or a Pell recipient and looks at how the TOP affects the diversity of the student body at the four stages of the enrollment funnel by employing a quasi-experimental research design (see below for details) and investigates the following two questions:

1. Does the test optional admission policy increase the probability that an applicant (accepted, enrolled or retained student) will be a minority group member?

To clarify the question, let us ask a statistical probability question: if you have an unlabeled folder of an applicant in front of you, does knowing that this applicant is a non-test submitter, increase the chance you can correctly predict whether s/he is a minority group member or not? From the institutional policy perspective, it can be rephrased: does allowing an applicant to opt out of the submission of test scores increase the probability that the applicant could be a member of minority groups?

2. Is the TOP impact on diversity the same at each stage of the enrollment funnel?

Research Design

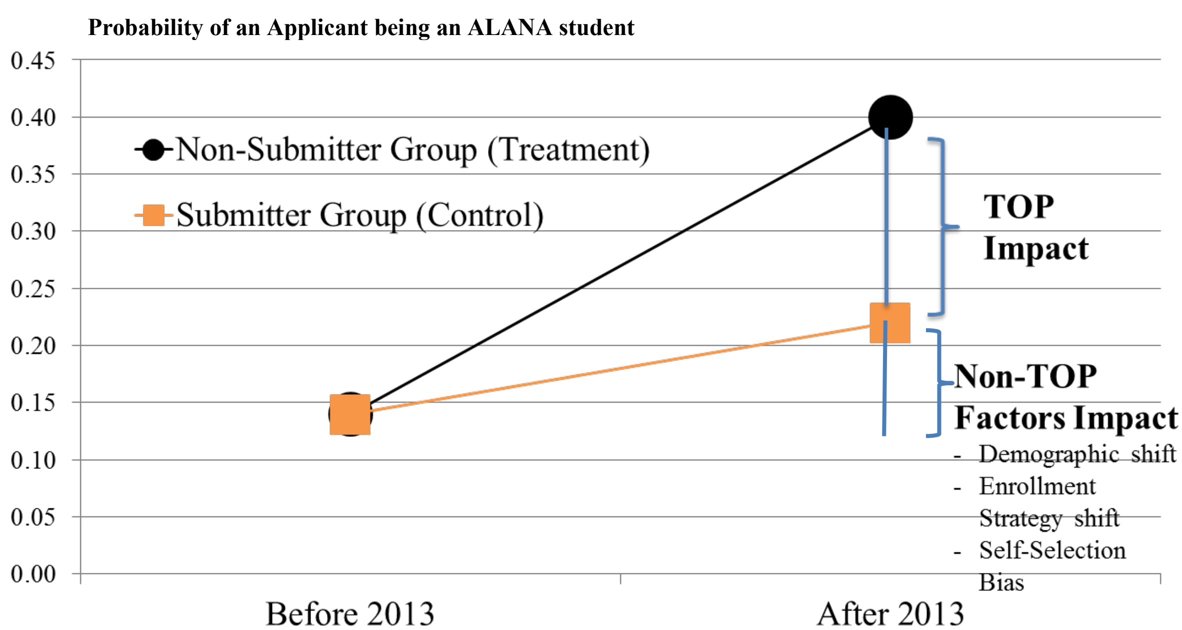
Ithaca College's TOP policy has been in effect for three years. The present study analyzes over 90,000 individual applicant records comparing Ithaca College's first three test-optional cohorts to the three cohorts prior to the test optional adoption.

This study employs a quasi-experimental research design with the DiD (Difference in Difference) analysis strategy. The students who did not submit standardized test scores for admission under the TOP form the treatment group; the students who submitted standardized test scores for admission form the control group. The control group in this study consists of two sub-groups: those who were required to submit test scores for admission before the College's TOP implementation and those who chose to submit test scores for admission after implementation of the new test optional policy in 2013. Continuing with the laboratory experiment analogy, the first control group before TOP is considered the “pure” control group. In contrast, the second control group, which is influenced by the presence of TOP, is considered the “contaminated” control group. Contaminations of control groups in classroom experiments have been discussed in-depth in the field of educational psychology (Craven, Mash, Debus and Jayasinghe, 2001; Doyle and Hickey, 2013). While discussing these studies in detail is beyond the scope of this study, the idea of contamination of control groups is very useful to the present study. We can argue that in comparison to the “pure” control group, the “contaminated” control group in this study carries certain bias factors such as self-selection biases; time-induced changes in the external environment (e.g. the racial composition change of high school graduates in Northeast); time-induced changes in the College’s enrollment strategies (e.g. introduction of the integrated marketing campaign and massive recruitment efforts specifically targeting minority communities) and other less observable bias factors.

By analyzing these three groups (the treatment group, the “pure” control group before the TOP, and the “contaminated” control group after the TOP), this study exploits the advantage of the DiD (Difference in Difference) analysis strategy. DiD analysis considers time-induced variation to control for potential observable or unobservable differences that exist between treatment and control groups, which might otherwise be attributed to the “treatment” itself (Gelman & Hill, 2006, Belasco et al., 2014). Our DiD analysis focuses on the differences observed between the treatment and the control groups after controlling for the shifts observed in the two control groups before and after the TOP adoption. This analysis strategy enables us to establish the causal relationship between TOP implementation and campus diversity as distinguished from the other plausible causal factors that may have affected the change in the dependent variable, (i.e. racial diversity on campus) in the absence of the TOP implementation (e.g. demographic shifts or recruitment strategy shifts).

The illustration below helps to understand our research design further. In measuring the probability of an applicant's being a minority member, the change observed in the test-submitter group before and after the 2013 TOP implementation represents the effects of various non-TOP factors such as self-selection biases, increase in minority high school graduates due to the demographic shift, or the increase in minority recruitment efforts discussed above. By controlling for such change, the study looks for a statistically significant positive effect in the non-test submitter group (the treatment group) which would indicate that the test optional policy did indeed increase the probability of an applicant being an ALANA (Afro-American, Latino/a, Asian or Native American) student.

Figure 1: Illustration of Research Design



Multivariate Statistical Tests

Logistic Regression is applied to examine whether the test optional policy increased the probability of an applicant being a minority member.

$$F(x) = 1/(1+ e^{-(\beta0 + \beta1*X1 + \ldots + \beta5*X5)}) + €$$

$$g(F(x)) = \ln(F(x)/1- F(x)) = \beta0 + \beta1*X1 + \ldots + \beta5*X5 + €$$

g(F(x)) is the logit function. The equation for g(F(x)) shows that the logit (natural logarithm of the odds) is equivalent to the multiple regression expression. Here,

F(x): 1 for ALANA (Afro-American, Latino/a, Asian and Native American) **Applicant and 0 for others; 1 for Pell Recipient and 0 for Non-Pell**

X1: HS GPA

X2: Family Contribution to Education (in $)

X3: NY State Resident or not

X4: 1 for before the TOP implementation in 2013; 0 for after 2013

X5: 1 for Non-submitters (Opted out Test Scores); 0 for Test-submitters

Our unpublished internal research reported that High School GPA, Family's ability to pay for education and New York State residency are the important variables that predict a correct ALANA membership of our applicants. Thus, X1, X2 and X3 are included in the model. If the test-submission status did indeed increase the probability of an applicant being a minority group member after controlling for the time-variant factor X4, β5 associated with the test-submission status should be significant in a positive direction. A standard DiD model usually includes one interaction term, which examines the interactive effects of time trends and pre-existing differences between treatment and control groups. Given that in the present study, the treatment group before 2013 was empty, the interaction term X4*X5 produces a statistical redundancy. As a result, only two main effects are included in this equation.

Inserting X4=0 and X5=0, we obtain the following equation for the test-submitters ("pure" control group) prior to 2013: $\mathbf{G(F(x)) = \beta 0 + \beta 1{*}X1 + \beta 2{*}X2 + \beta 3{*}X3 + error}$

With X4=1 and X5=0, the following equation is derived for the "contaminated" control group after 2013: $\mathbf{G(F(x)) = (\beta 0+\beta 4) + \beta 1{*}X1 + \beta 2{*}X2 + \beta 3{*}X3 + error}$

Lastly, with X4=1 and X5=1, the following equation is obtained for the non-submitter (treatment) group after 2013: $\mathbf{G\,(F(x)) = (\beta 0+\beta 4+\beta 5) + \beta 1{*}X1 + \beta 2{*}X2 + \beta 3{*}X3 + error}$

The present study would find a statistical significance associated with β5 in a positive direction if the test-submission status did indeed increase the probability of an applicant being a minority community member after controlling for the time-variant and other bias factors expressed in X4.

Descriptive Results

Descriptive statistics of the two dependent variables and the five independent variables are presented in Tables 1 to 4. These basic statistics are presented by test-submission status. Correlation analyses of those variables are also presented in Tables 5 through 9.

About 10% of the applicant population did not submit high school cumulative GPA data, but only 1% of the admitted, enrolled and retained population had high school GPA missing data. The socio-economic advantages of the test-submitters in comparison to the non-test submitters' are revealed by Tables 1 through 4. In the post-2013 period, the average family contribution of the test-submitters was more than their counterpart's by \$5,600, \$6,600, \$4,700 and \$4,300

Table 1
Descriptive Analysis of Variables by Test Submission Status
Applicant Population

Type of Variables		Dependent		Covariates			Dichotomous	
		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
Test-Submitters (Before 2013)	N	40440	40440	40440	36305	40440	40440	40440
	N of category=1	9172	4513	16271	NA	NA	0	0
	Mean	.2268	.1116	.4023	3.2827	$29,433	.0	.0
	Std. Dev	.4188	.3149	.4904	.5301	$20,830	.0	.0
	Minimum	0	0	0	1.00	$0	0	0
	Maximum	1	1	1	4.48	$54,717	0	0
Test-Submitters (After 2013)	N	37564	37564	37564	34790	37564	37564	37564
	N of category=1	9666	3788	14576	NA	NA	37564	0
	Mean	.2573	.1008	.3880	3.2986	$36,667	1	.0
	Std. Dev	.4372	.3011	.4873	.4994	$22,057	.0	.0
	Minimum	0	0	0	1.00	$0	1	0
	Maximum	1	1	1	4.59	$61,258	1	0
Non-Test Submitters (After 2013)	N	12820	12820	12820	11681	12820	12820	12820
	N of category=1	5097	2184	6115	NA	NA	12820	12820
	Mean	.3976	.1704	.4770	3.1636	$31,062	1	1
	Std. Dev	.4894	.3760	.4995	.5068	$24,260	.0	.0
	Minimum	0	0	0	.99	$0	1	1
	Maximum	1	1	1	4.32	$60,585	1	1
Total	N	90824	90824	90824	82776	90824	90824	90824
	N of category=1	23935	10485	36962	NA	NA	50384	12820
	Mean	.2635	.1154	.4070	3.2726	$32,655	.5547	.1412
	Std. Dev	.4406	.3196	.4913	.5160	$22,116	.4970	.3482
	Minimum	0	0	0	.99	$0	0	0
	Maximum	1	1	1	4.59	$61,258	1	1

Table 2
Descriptive Analysis of Variables by Test Submission Status
Admitted Population

Type of Variables		Dependent		Covariates			Dichotomous	
		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
Test-Submitters (Before 2013)	N	27222	27222	27222	26913	27222	27222	27222
	N of category=1	5116	4503	11041	NA	NA	0	0
	Mean	.1879	.1654	.4056	3.3794	$31,433	.0	.0
	Std. Dev	.3907	.3716	.4910	.4767	$19,016	.0	.0
	Minimum	0	0	0	1.00	$0	0	0
	Maximum	1	1	1	4.48	$54,717	0	0
Test-Submitters (After 2013)	N	24633	24633	24633	24470	24633	24633	24633
	N of category=1	5380	3784	9623	NA	NA	24633	0
	Mean	.2184	.1536	.3907	3.3529	$36,517	1	.0
	Std. Dev	.4132	.3606	.4879	.4433	$21,233	.0	.0
	Minimum	0	0	0	1.00	$0	1	0
	Maximum	1	1	1	4.59	$61,258	1	0
Non-Test Submitters (After 2013)	N	7631	7631	7631	7581	7631	7631	7631
	N of category=1	2700	2180	3567	NA	NA	7631	7631
	Mean	.3538	.2857	.4674	3.2657	$29,954	1	1
	Std. Dev	.4782	.4518	.4990	.4333	$23,353	.0	.0
	Minimum	0	0	0	1.63	$0	1	1
	Maximum	1	1	1	4.08	$60,585	1	1
Total	N	59486	59486	59486	58964	59486	59486	59486
	N of category=1	13196	10467	24231	NA	NA	32264	7631
	Mean	.2218	.1760	.4073	3.3538	$33,349	.5424	.1283
	Std. Dev	.4155	.3808	.4913	.4590	$20,723	.4982	.3344
	Minimum	0	0	0	1.00	$0	0	0
	Maximum	1	1	1	4.59	$61,258	1	1

Table 3

Descriptive Analysis of Variables by Test Submission Status

Enrolled Population

Type of Variables		Dependent		Covariates			Dichotomous	
		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
Test-Submitters (Before 2013)	N	4893	4893	4893	4803	4893	4893	4893
	N of category=1	859	1012	2072	NA	NA	0	0
	Mean	.1756	.2068	.4235	3.3612	$28,392	.0	.0
	Std. Dev	.3805	.4051	.4942	.5130	$18,761	.0	.0
	Minimum	0	0	0	1.00	$0	0	0
	Maximum	1	1	1	4.00	$54,717	0	0
Test-Submitters (After 2013)	N	3789	3789	3789	3772	3789	3789	3789
	N of category=1	715	666	1581	NA	NA	3789	0
	Mean	.1887	.1758	.4173	3.3273	$33,657	1	.0
	Std. Dev	.3913	.3807	.4932	.4759	$20,582	.0	.0
	Minimum	0	0	0	1.52	$0	1	0
	Maximum	1	1	1	4.00	$61,258	1	0
Non-Test Submitters (After 2013)	N	1453	1453	1453	1443	1453	1453	1453
	N of category=1	450	433	740	NA	NA	1453	1453
	Mean	.3097	.2980	.5093	3.2225	$29,002	1	1
	Std. Dev	.4625	.4575	.5001	.4632	$22,390	.0	.0
	Minimum	0	0	0	1.63	$0	1	1
	Maximum	1	1	1	4.04	$60,585	1	1
Total	N	10135	10135	10135	10018	10135	10135	10135
	N of category=1	2024	2111	4393	NA	NA	5242	1453
	Mean	.1997	.2083	.4334	3.3285	$30,448	.5172	.1434
	Std. Dev	.3998	.4061	.4956	.4944	$20,156	.4997	.3505
	Minimum	0	0	0	1.00	$0	0	0
	Maximum	1	1	1	4.04	$61,258	1	1

Table 4

Descriptive Analysis of Variables by Test Submission Status

Retained Population

Type of Variables		Dependent		Covariates			Dichotomous	
		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
Test-Submitters (Before 2013)	N	4117	4117	4117	4056	4117	4117	4117
	N of category=1	693	801	1744	NA	NA	0	0
	Mean	.1683	.1946	.4236	3.3876	$28,958	.0	.0
	Std. Dev	.3742	.3959	.4942	.5060	$18,652	.0	.0
	Minimum	0	0	0	1.79	$0	0	0
	Maximum	1	1	1	4.00	$54,717	0	0
Test-Submitters (After 2013)	N	2091	2091	2091	2082	2091	2091	2091
	N of category=1	367	334	855	NA	NA	2091	0
	Mean	.1755	.1597	.4089	3.3602	$33,784	1	.0
	Std. Dev	.3805	.3664	.4917	.4659	$19,982	.0	.0
	Minimum	0	0	0	1.52	$0	1	0
	Maximum	1	1	1	4.00	$58,902	1	0
Non-Test Submitters (After 2013)	N	770	770	770	766	770	770	770
	N of category=1	234	227	404	NA	NA	770	770
	Mean	.3039	.2948	.5247	3.2505	$29,497	1	1
	Std. Dev	.4602	.4563	.4997	.4492	$21,829	.0	.0
	Minimum	0	0	0	1.95	$0	1	1
	Maximum	1	1	1	4.00	$59,602	1	1
Total	N	6978	6978	6978	6904	6978	6978	6978
	N of category=1	1294	1362	3003	NA	NA	2861	770
	Mean	.1854	.1952	.4304	3.3641	$30,463	.4100	.1103
	Std. Dev	.3887	.3964	.4952	.4899	$19,548	.4919	.3133
	Minimum	0	0	0	1.52	$0	0	0
	Maximum	1	1	1	4.00	$59,602	1	1

Table 5
Correlation Analysis
Applicant Population

		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
ALANA	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	90824						
Pell	Pearson Correlation	.189**	1					
	Sig. (2-tailed)	0.000						
	N	90824	90824					
NY	Pearson Correlation	.153**	.112**	1				
	Sig. (2-tailed)	0.000	.000					
	N	90824	90824	90824				
HS_GPA	Pearson Correlation	-.139**	.078**	.093**	1			
	Sig. (2-tailed)	0.000	.000	.000				
	N	82776	82776	82776	82776			
Family Contribution	Pearson Correlation	-.246**	-.432**	-.179**	-.035**	1		
	Sig. (2-tailed)	0.000	0.000	0.000	.000			
	N	90824	90824	90824	82776	90824		
After 2013	Pearson Correlation	.075**	.011**	.008*	-.017**	.131**	1	
	Sig. (2-tailed)	.000	.001	.011	.000	0.000		
	N	90824	90824	90824	82776	90824	90824	
Non-Test Submitters (Treatment)	Pearson Correlation	.123**	.070**	.058**	-.086**	-.029**	.363**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	0.000	
	N	90824	90824	90824	82776	90824	90824	90824

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 6
Correlation Analysis
Admits Population

		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
ALANA	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	59486						
Pell	Pearson Correlation	.303**	1					
	Sig. (2-tailed)	0.000						
	N	59486	59486					
NY	Pearson Correlation	.102**	.142**	1				
	Sig. (2-tailed)	0.000	.000					
	N	59486	59486	59486				
HS_GPA	Pearson Correlation	-.059**	.025**	.201**	1			
	Sig. (2-tailed)	0.000	.000	.000				
	N	58964	58964	58964	58964			
Family Contribution	Pearson Correlation	-.262**	-.606**	-.180**	-.127**	1		
	Sig. (2-tailed)	0.000	0.000	0.000	.000			
	N	59486	59486	59486	58964	59486		
After 2013	Pearson Correlation	.075**	.025**	0.003	-.051**	.085**	1	
	Sig. (2-tailed)	.000	.000	.425	.000	0.000		
	N	59486	59486	59486	58964	59486	59486	
Non-Test Submitters (Treatment)	Pearson Correlation	.122**	.111**	.047**	-.074**	-.063**	.352**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	0.000	
	N	59486	59486	59486	58964	59486	59486	59486

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 7
Correlation Analysis
Enrolled Population

		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
ALANA	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	10135						
Pell	Pearson Correlation	**.295****	1					
	Sig. (2-tailed)	0.000						
	N	10135	10135					
NY	Pearson Correlation	**.082****	**.170****	1				
	Sig. (2-tailed)	0.000	.000					
	N	10135	10135	10135				
HS_GPA	Pearson Correlation	**-.081****	**.032****	**.192****	1			
	Sig. (2-tailed)	0.000	.002	.000				
	N	10018	10018	10018	10018			
Family Contribution	Pearson Correlation	**-.266****	**-.609****	**-.197****	**-.156****	1		
	Sig. (2-tailed)	0.000	0.000	0.000	.000			
	N	10135	10135	10135	10018	10135		
After 2013	Pearson Correlation	**.058****	**0.003**	**.019***	**-.064****	**.099****	1	
	Sig. (2-tailed)	.000	.726	.050	.000	0.000		
	N	10135	10135	10135	10018	10135	10135	
Non-Test Submitters (Treatment)	Pearson Correlation	**.113****	**.090****	**.063****	**-.088****	**-.029****	**.395****	1
	Sig. (2-tailed)	.000	.000	.000	.000	.003	0.000	
	N	10135	10135	10135	10018	10135	10135	10135

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 8
Correlation Analysis
Retained Population

		ALANA	Pell	NY	HS_GPA	Family Contribution	After 2013	Non-Test Submitters (Treatment)
ALANA	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	6978						
Pell	Pearson Correlation	**.295****	1					
	Sig. (2-tailed)	0.000						
	N	6978	6978					
NY	Pearson Correlation	**.075****	**.176****	1				
	Sig. (2-tailed)	0.000	.000					
	N	6978	6978	6978				
HS_GPA	Pearson Correlation	**-.065****	**.032****	**.202****	1			
	Sig. (2-tailed)	0.000	.009	.000				
	N	6904	6904	6904	6904			
Family Contribution	Pearson Correlation	**-.271****	**-.598****	**-.204****	**-.160****	1		
	Sig. (2-tailed)	0.000	0.000	0.000	.000			
	N	6978	6978	6978	6904	6978		
After 2013	Pearson Correlation	**.053****	**0.002**	**0.016**	**-.057****	**.092****	1	
	Sig. (2-tailed)	.000	.874	.172	.000	0.000		
	N	6978	6978	6978	6904	6978	6978	
Non-Test Submitters (Treatment)	Pearson Correlation	**.107****	**.089****	**.067****	**-.082****	**-0.0174**	**.422****	1
	Sig. (2-tailed)	.000	.000	.000	.000	.146	0.000	
	N	6978	6978	6978	6904	6978	6978	6978

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

respectively at the applied, admitted, enrolled and retained stage of the enrollment funnel. Similarly, the average of high school cumulative GPA among the test-submitters was higher than the average of the non-test submitters by 0.14, 0.09, 0.11 and 0.11 respectively at the applied, admitted, enrolled and retained stage of the enrollment funnel.

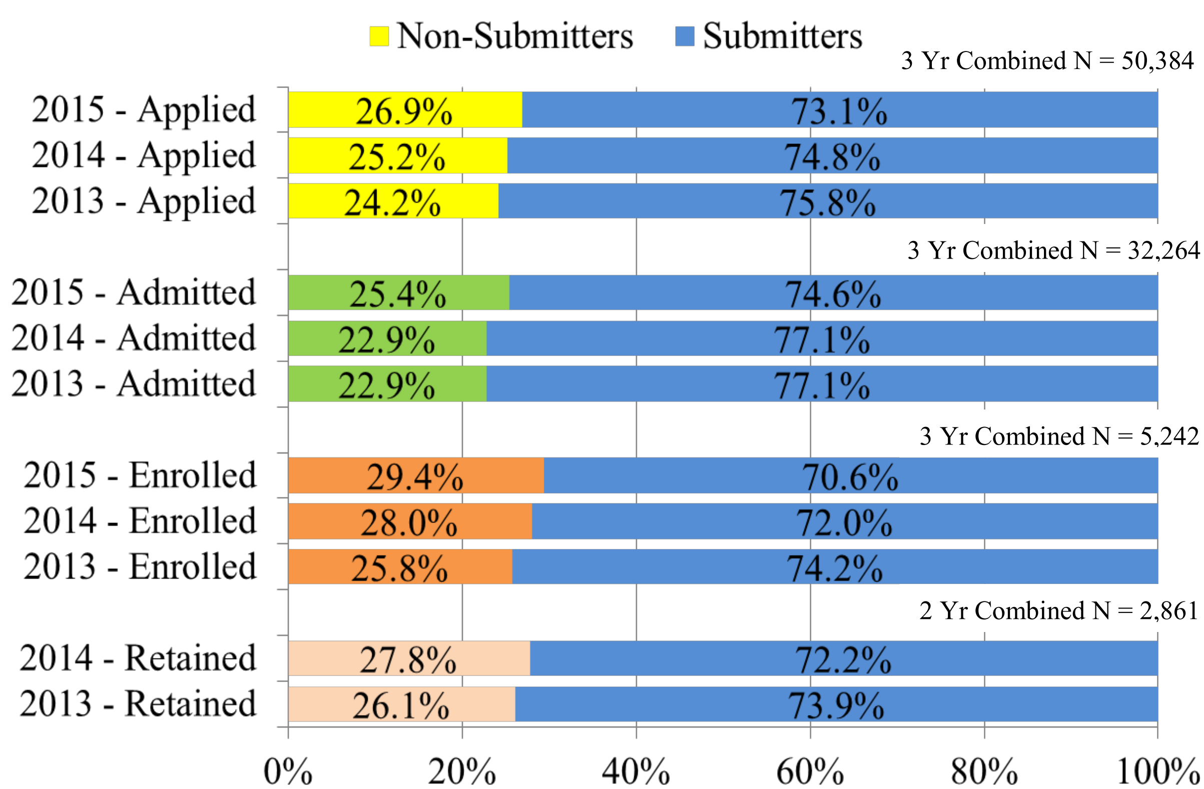
Three dichotomous variables, such as New York State residency, a before-or-after-2013 indicator, and test-submitter status, are significantly correlated positively with the ALANA status at each stage of the enrollment funnel. Two other variables (high school cumulative GPA and amount of family contributions to education calculated by the institutional methodology) are negatively correlated with ALANA status at each stage of the enrollment funnel as shown in Tables 5 to 8.

The correlation statistics associated with the Pell recipient status are similar to the above described correlation analysis on the ALANA variable except for high school cumulative GPA, which is significantly correlated with Pell status in the positive direction, indicating that the Pell recipients who applied, admitted, enrolled and retained at Ithaca, are better academic achievers at high school than the non-Pell recipient population.

In Figure 2, the colored bars indicate the percentage of non-test submitters of the population at each stage of the funnel from 2013 to 2015. Several important finding are shown. First, the percentage of non-test submitters has steadily increased over three years; and second, the proportion of non-test submitters was smaller at the admitted stage than at the application stage, implying non-submitters have had lower admit rates. But the proportion of non-submitters increased at the enrollment stage which indicates a higher yield rate for the TOP group. One explanation for this is that at Ithaca College, all accepted applicants including test optional students are considered for merit awards based on composite scores of four academic measures.

If a standardized test score is not submitted, the average of the remaining three measures is used to calculate a composite score for merit award consideration. Given that many of our competing schools require standardized test scores for merit scholarship consideration, Ithaca's merit awarding policy might be helping the non-test submitters enroll at the higher rate than the test-submitters. Lastly, the proportion of the non-test submitters changed very little at the third semester retention stage, which implies the non-test submitters were retained as well as the test-submitters, as indicated Hiss and Franks (2014).

Figure 2: Non-Test Submitters % by Funnel



Figures 3 - 6 show the ALANA (minority) percentage of non-test submitters in colored bars compared to the ALANA percentage of test submitters in blue bars at each of the four stages of the enrollment funnel. Figure 3 is for the applicant population. It clearly shows that the ALANA representation is higher by thirteen to fifteen percentage points among non-test submitters than among test submitters. We can also observe that there was a steady upward increase in the ALANA percent among test-submitters over six years.

Figure 4 shows the same data for the admitted population. The ALANA percentages are slightly lower than in the applied population for both submitter and non-submitter groups, but we can still observe the higher ALANA representation in the non-submitter group by approximately 13 percent. Figures 5 and 6 are for the enrolled population and the retained population respectively. The ALANA representations declined again in comparison to the applied and admitted populations. The ALANA percentage difference between submitter and non-submitter groups also shrank to about ten percentage points in 2015.

When we observe Figures 3 through 6 sequentially, it is clear that the College loses the ALANA representation at each stage of the enrollment funnel. The positive news here is that the TOP seems to boost the ALANA representation at each stage. As discussed earlier, logistic regression under the well-crafted quasi-experimental design confirms this statement.

Multivariate Analysis Results

Tables 9 indicates the logistic regression output for the applicant population, which displays the variables in the equation, beta coefficients, standard error of coefficients, Wald statistics, degree of freedom, statistical significance level associated with betas, and odds ratios. The most important question here is whether X5, the participation in the test optional, contributed to the

Figure 3: ALANA % of Applicants
Test-Submitters vs. Non Test-Submitters

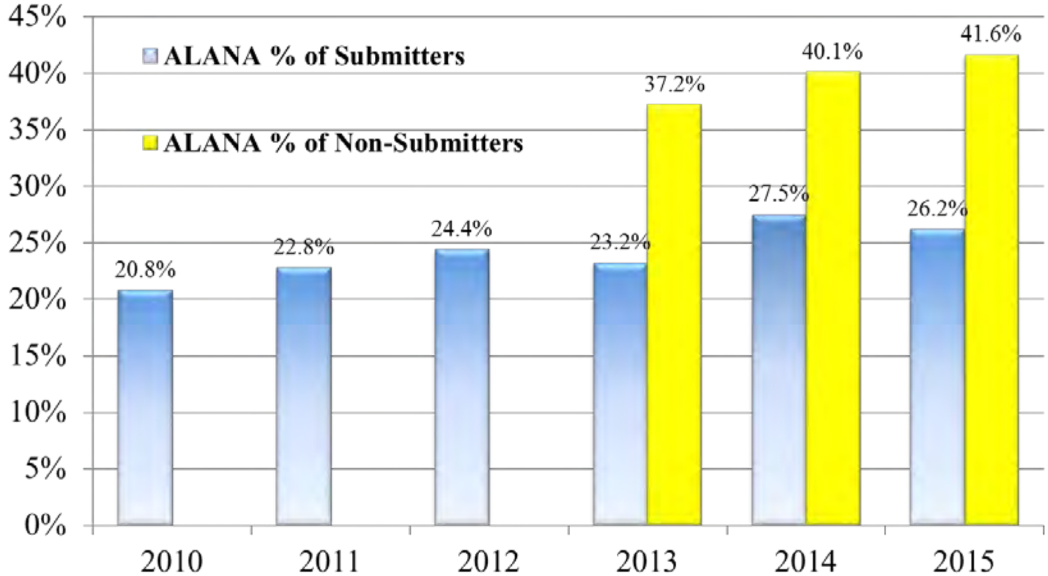


Figure 4: ALANA % of Admitted
Test-Submitters vs. Non Test-Submitters

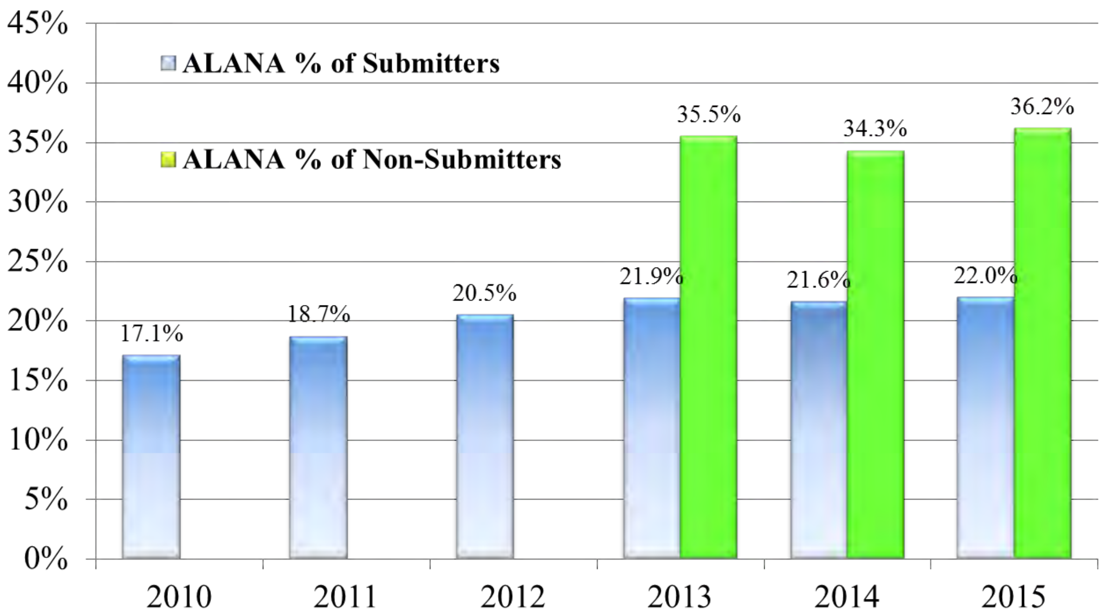


Figure 5: ALANA % of Enrolled Students
Test-Submitters vs. Non Test-Submitters

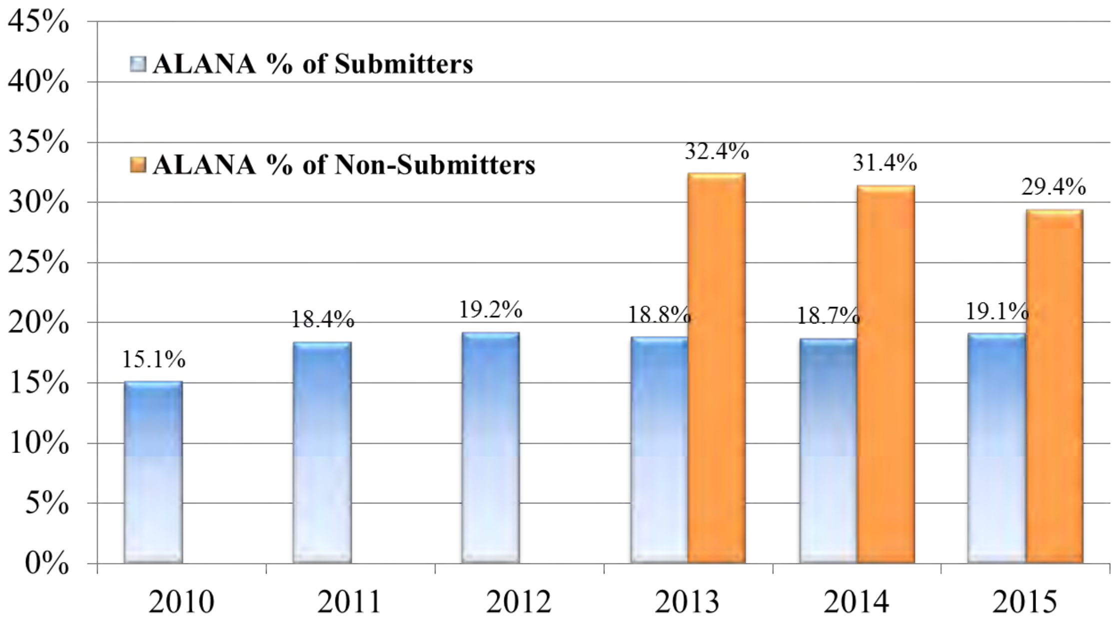
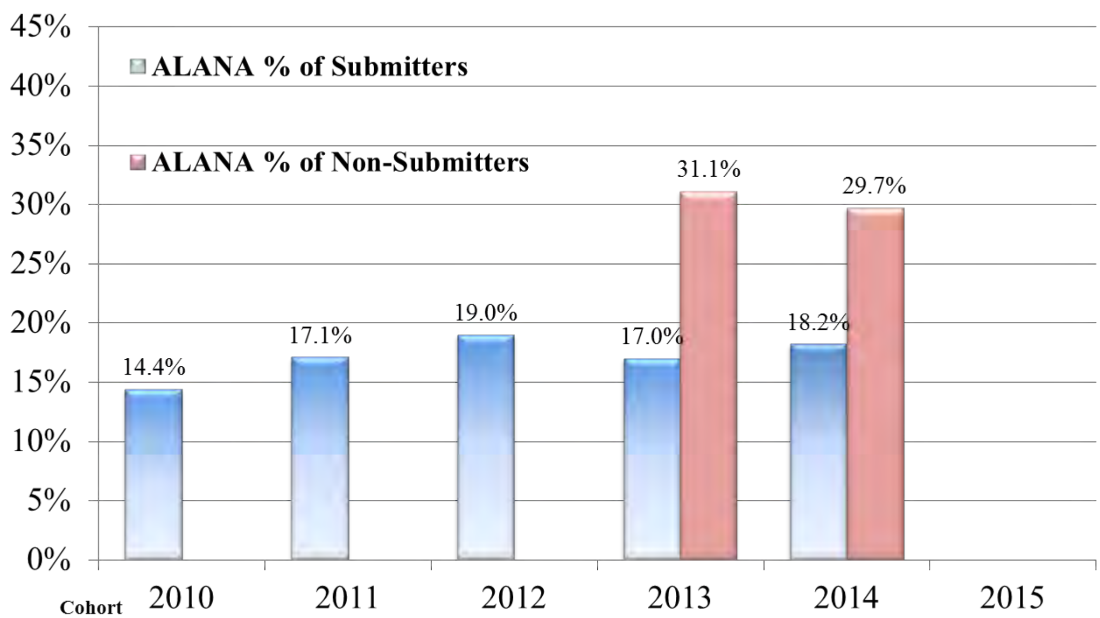


Figure 6: ALANA % of Retained at 3rd Semester
Test-Submitters vs. Non Test-Submitters



increase in the probability of an applicant being an ALANA member, after controlling for the effects associated with the change before and after the 2013 TOP implementation. As discussed earlier, the change observed in the test-submitter groups before and after the 2013 TOP implementation represents the effects related to various non-TOP factors such as self-selection biases, the increase in minority high school graduates due to the demographic shift, and the increase in minority recruitment efforts. If the test optional policy did indeed increase the probability of an applicant being an ALANA student by controlling for the non-TOP effects expressed in X4, the study should observe a statistically significant, positive beta coefficient ($\beta 5$) associated with treatment (the non-test submitter) status.

As shown in Table 9, $\beta 5$ is highly significant in the positive direction confirming the positive impact of the test optional policy on increasing the probability of an applicant being a minority student. Interestingly, $\beta 4$ was significant in the positive direction indicating that time induced non-test optional factors also contributed to the increase in the probability. High school GPA, amount of family contribution to education and New York residency are powerful variables to predict an ALANA membership as indicated by previous research.

Similar findings are revealed for the admitted population as presented in Table 10. Again, $\beta 5$ is statistically significant in the positive direction indicating that the probability of an accepted applicant being an ALANA member has increased among the non-test submitters. As mentioned above, the two landmark studies discussed earlier did not examine how the institution's admission decisions affected enrolled diversity under the test optional policy. In Table 10, the present study presents the first evidence to confirm the positive impact of the TOP on the institution's admission process. Notice that stating the probability of an accepted applicant being an ALANA member has increased under the TOP, is not the same as saying that the admit rate of

ALANA students has increased under the TOP. The former asks about the chance of an admitted applicant being an ALANA student in the admitted population. In contrast, the latter asks about the ALANA student' chance of being admitted, by looking at the ratio of the accepted to the applied. The present study exclusively deals with the first question. The next two tables (Tables 11 and 12) show similar findings, that is, the positive contribution of the TOP to increase the ALANA representation at the enrolled and retained stages of the funnel.

The logistic regression analyses are repeated using Pell Grant recipient status (1 for Pell recipients and 0 for others) as a dependent variable. The regression results are shown in Tables 13 – 16. $\beta 5$ in each table is highly significant in the positive direction ($p<.000$ at application, admission and retention stages and $p<.10$ at enrollment stage), confirming the positive impact of the test optional policy on increasing the probability of a student being a Pell recipient.

Summary and Conclusion

Currently, over 850 institutions, including several national universities such as Wake Forest and George Washington, have adopted a "test-optional policy" (TOP). This policy advocates the increase in campus diversity by removing the barriers against various minority groups often presented by standardized testing. In-depth research on the impact of the TOP on campus diversity is, however, still in its early stages. The present study looks at each stage of the enrollment funnel and asks "Does the test optional admission policy increase the probability that a student will be a minority group member?" This study is an effort to provide an institutionally-specific research example to other institutions so that more research results can be compiled and shared to advance our understanding of the impact of the test optional admission policy.

Table 9

Logistic Regression Result: Applicants

N=82,222	Dependent Var: ALANA = 1 Non-ALANA= 0					
	B	S.E.	Wald	df	Sig.	Exp(B)
X1 NY_STATE	.547	.017	998.476	1	.000	1.728
X2 HS_GPA	-.710	.016	1855.330	1	.000	.492
X3 Family Contribution	-.274	.004	4749.475	1	.000	.760
X4 After2013	.412	.019	473.602	1	.000	1.510
X5 Test_Optional	.412	.019	473.602	1	.000	1.510
Constant	1.503	.056	721.075	1	.000	4.494
Nagelkerke R-sqr =	0.165 (<.000)					

Table 10

Logistic Regression Result: Admits

N=58,676	Dependent Var: ALANA = 1 Non-ALANA= 0					
	B	S.E.	Wald	df	Sig.	Exp(B)
X1 NY_STATE	.362	.022	278.909	1	.000	1.436
X2 HS_GPA	-.547	.023	542.732	1	.000	.579
X3 Family Contribution	-.314	.005	3587.601	1	.000	.731
X4 After2013	.329	.023	203.991	1	.000	1.389
X5 Test_Optional	.441	.031	208.610	1	.000	1.555
Constant	1.098	.082	180.308	1	.000	2.997
Nagelkerke R-sqr =	0.139 (<.000)					

Table 11

Logistic Regression Result: Enrolled

N=10,011	Dependent Var: ALANA = 1 Non-ALANA= 0					
	B	S.E.	Wald	df	Sig.	Exp(B)
X1 NY_STATE	.231	.055	17.818	1	.000	1.260
X2 HS_GPA	-.662	.055	145.346	1	.000	.516
X3 Family Contribution	-.375	.015	658.655	1	.000	.687
X4 After2013	.243	.059	16.849	1	.000	1.274
X5 Test_Optional	.442	.076	34.133	1	.000	1.557
Constant	1.475	.191	59.462	1	.000	4.371
Nagelkerke R-sqr =	0.151 (<.000)					

Table 12

Logistic Regression Result: Retained

N=6,882	Dependent Var: ALANA = 1 Non-ALANA= 0					
	B	S.E.	Wald	df	Sig.	Exp(B)
X1 NY_STATE	.162	.068	5.657	1	.017	1.176
X2 HS_GPA	-.592	.069	74.346	1	.000	.553
X3 Family Contribution	-.406	.019	467.945	1	.000	.666
X4 After2013	.202	.075	7.269	1	.007	1.224
X5 Test_Optional	.529	.105	25.526	1	.000	1.698
Constant	1.324	.240	30.399	1	.000	3.760
Nagelkerke R-sqr =	0.152 (<.000)					

Table 13

Logistic Regression Result: Applicants

N=82,222 **Dependent Var: Pell Recipients = 1 Not Pell Recipients= 0**

		B	S.E.	Wald	df	Sig.	Exp(B)
X1	NY_STATE	.055	.025	4.696	1	.030	1.056
X2	HS_GPA	.812	.025	1044.748	1	.000	2.252
X3	Family Contribution	-1.208	.014	6990.552	1	.000	.299
X4	After2013	.270	.028	93.845	1	.000	1.309
X5	Test_Optional	.211	.036	34.555	1	.000	1.235
	Constant	-2.874	.085	1139.924	1	.000	0.056

Nagelkerke R-sqr = 0.449 (<.000)

Table 14

Logistic Regression Result: Admits

N=58,676 **Dependent Var: Pell Recipients = 1 Not Pell Recipients= 0**

		B	S.E.	Wald	df	Sig.	Exp(B)
X1	NY_STATE	.167	.032	27.212	1	.000	1.181
X2	HS_GPA	-.242	.036	46.073	1	.000	.785
X3	Family Contribution	-1.888	.022	7353.818	1	.000	.151
X4	After2013	.173	.034	25.188	1	.000	1.188
X5	Test_Optional	.197	.047	17.255	1	.000	1.218
	Constant	2.190	.125	305.720	1	.000	8.934

Nagelkerke R-sqr = 0.650 (<.000)

Table 15

Logistic Regression Result: Enrolled

N=10,011 **Dependent Var: Pell Recipients = 1 Not Pell Recipients= 0**

		B	S.E.	Wald	df	Sig.	Exp(B)
X1	NY_STATE	.286	.074	14.745	1	.000	1.331
X2	HS_GPA	-.328	.078	17.501	1	.000	.721
X3	Family Contribution	-2.021	.052	1498.978	1	.000	.133
X4	After2013	.199	.081	6.047	1	.014	1.220
X5	Test_Optional	.197	.113	3.064	1	.080	1.218
	Constant	2.718	.277	96.168	1	.000	15.155

Nagelkerke R-sqr = 0.661 (<.000)

Table 16

Logistic Regression Result: Retained

N=6,882 **Dependent Var: Pell Recipients = 1 Not Pell Recipients= 0**

		B	S.E.	Wald	df	Sig.	Exp(B)
X1	NY_STATE	.321	.090	12.566	1	.000	1.378
X2	HS_GPA	-.303	.096	10.072	1	.002	.738
X3	Family Contribution	-1.972	.062	1000.631	1	.000	.139
X4	After2013	.148	.102	2.081	1	.149	1.159
X5	Test_Optional	.443	.155	8.136	1	.004	1.557
	Constant	2.541	.338	56.456	1	.000	12.692

Nagelkerke R-sqr = 0.646 (<.000)

Ithaca College, a mid-sized four-year residential comprehensive college in central New York, implemented the policy in 2012 for the admission applications of the fall 2013 entering cohort. This study analyzed over 90,000 individual applicant records from the three test-optional cohorts and the three cohorts prior to the implementation of TOP at Ithaca College. The study is a first attempt to reveal the insights of how the TOP affected the diversity of the student body at four stages of the enrollment funnel: application, admission, enrollment, and retention. A minority group member was defined as a member of a racial minority or a Pell recipient.

The study employed a quasi-experimental research design with the DiD (Difference in Difference) analysis strategy. The applicants who did not submit test scores for admission under Ithaca's test optional policy formed the treatment group. In contrast, the control group in this study consisted of two sub-groups: those who were required to submit test scores for admission before the College's TOP implementation ("pure" control group) and those who chose to submit test scores for admission after implementation of the new test optional policy in 2013 ("contaminated" control group). In comparison to the "pure" control group, the "contaminated" control group in the study carries certain bias factors such as self-selection biases; time-induced changes in the external environment (e.g. racial composition change of the high school graduates in Northeast); time-induced changes in the College's enrollment strategies (e.g. massive recruitment efforts specifically targeting minority communities) and other biases. Our DiD analysis has focused on the differences observed between the treatment and the control groups after controlling for the shifts observed in the two control groups before and after the TOP adoption. This analysis strategy has enabled us to establish the causal relationship between TOP implementation and campus diversity as distinguished from other plausible causal factors that may have affected the change in diversity in the absence of the TOP implementation (e.g. demographic shifts or recruitment strategy changes).

Logistic regression analysis under this quasi-experimental design has revealed that the beta coefficient (β5) associated with the treatment (the non-test submitter) status was statistically significant in the positive direction at each stage of the enrollment funnel after the non-TOP effects were appropriately controlled for. The results confirmed the positive impact of the test optional policy on the increase in the probability of a candidate representing a minority group.

It is true that this conclusion is drawn based on only one institution's data. This one-school setting coupled with the well-constructed research design did indeed enable us to distinguish the test optional effect from other plausible causal explanations. This study also revealed insights about how the TOP affected the diversity of the student body at application, admission, enrollment, and retention stages of the enrollment funnel, which has never been investigated before. In conclusion, the present study has provided valuable information to complement the findings of the large national landmark studies and suggested a number of important topics for future research (e.g. the impact of TOP on financial aid). The author hopes that other institutions will use this study as an institutionally-specific research example to advance our understanding of the impact of the test optional admission policy on the U.S. higher education landscape.

References

Atkinson, R. C. (2001). *Standardized Tests and Access to American Universities*, The 2001 Robert H. Atwell Distinguished Lecture, delivered at the Eighty-third Annual Meeting of the American Council on Education, Washington, D.C., February 18, 2001, retrieved from http://works.bepress.com/richard_atkinson/36/

Belasco, A. S., Rosinger, K. O., and Hearn J. C. (2014). "*The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boom for Equity or Something Else?*"

Educational Evaluation and Policy Analysis, AERA, Month 201X, Vol. XX, No. X, pp. 1-18.

Copeland, T. (2009) "*Introducing the Enrollment Funnel -A Valuable Tool for Professional and Continuing Education Outreach*," retrieved from http://www.enrollmentmarketing.org/research/WP-Introducing-CE-Enrollment-Funnel-Tim-Copeland.pdf

Cornwell, C., Mustard, D. B. and Van Parys, J. (2012). "The New SAT and Academic Performance at a Flagship Public University," in *SAT Wars: The Case for Test-Optional College Admissions*, edited by J. A. Soares, Teachers College, Columbia University, New York and London.

Craven, R. G., Mash. H. W., Debus, R. L., and Jayasinghe, U. (2001). "*Diffusion Effects: Control Group Contamination Threats to the Validity of Teacher-Administered Interventions*." Journal of Educational Psychology, Vol. 93, No. 3, pp. 639-645.

Doyle, O., and Hickey, C., (2013). "The challenges of contamination in evaluations of childhood interventions," SAGE Journals: Evaluation, pp. 183-194, retrieved from http://evi.sagepub.com/content/19/2/183.full.pdf+html

Epstein J. P., (2009). Behind the SAT-Optional Movement: Context and Controversy, retrieved from http://www.maguireassoc.com/resource/documents/SAT_optional_article-NACAC_Journal_College_Admission_J.Epstein_7-09.pdf

FairTest: The National Center for Fair and Open Testing. (2015). "850 Colleges, Universities Are Test-Optional, Including 195 Top-Tier Schools," retrieved from http://www.fairtest.org/850-colleges-universities-are-testoptional-includi

Gelman, A., and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York, NY: Cambridge University Press.

Hiss, W. C., and Franks, V. W., (2014). "*Defining Promise: Optional Standardized Testing Policies in American College and University Admissions*" retrieved from http://www.nacacnet.org/research/research-data/nacac-research/Documents/DefiningPromise.pdf

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., and Barbuti, S. M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average*, College Board Research Report: No. 2008-5, The college, Board, New York.

Mulugetta, Y. (2013), "*Going Test Optional: Gathering Evidence and Making the Decision at Ithaca College*," in the 40th NEAIR Conference Proceeding.

National Association for College Admissions Counseling (NACAC). (2008). *Report of the Commission on the Use of Standardized Tests in Undergraduate Admissions*. Arlington, VA: Author.

Rask, K. and Tiefenthaler, J. (2012). "The SAT as a Predictor of Success at a Liberal Arts College," in *SAT Wars: The Case for Test-Optional College Admissions*, edited by J. A. Soares, Teachers College, Columbia University, New York and London.

Rochon, T. (2013), *The Case Against the SAT*, USNEWS.COM/Opinion, 9/06/2013 retrieved from http://www.usnews.com/opinion/articles/2013/09/06/why-the-sats-shouldnt-be-a-factor-in-college-admissions

Rosner, J. (2012). "The SAT: Quantifying the Unfairness Behind the Bubbles," in *SAT Wars: The Case for Test-Optional College Admissions*, edited by J. A. Soares, Teachers College, Columbia University, New York and London.

Simon, C. C. (2015). "The Test-Optional Surge," in Education Life, New York Times, 10/28/2015, retrieved from http://www.nytimes.com/2015/11/01/education/edlife/the-test-optional-surge.html?_r=0

Wonnell, T., Rothstein, C. M. and Latting, J. (2012). "Predictor of Success at Highly Selective Private Research University," in *SAT Wars: The Case for Test-Optional College Admissions*, edited by J. A. Soares, Teachers College, Columbia University, New York and London.

Zwick, R., ed. (2004). *Rethinking the SAT: The Future of Standardized Testing in University Admissions*, New York, RoutledgeFalmer.

Zwick, R., (2007). "College Admission Testing," National Association for College Admission Counseling, Arlington, VA.

ZOLTAR SPEAKS: WILL YOU COMPLETE YOUR ONLINE COURSE?

Joseph King, Senior Evaluation Specialist
Katherine Ostroot, Institutional Research Specialist
Joanna Ostroot, Student Assistant
Office of Decision Support
State University of New York Empire State College

Abstract

The majority of students at Empire State College are at-risk students, many of whom pursue online education. The college is not currently assessing the attributes, prior educational history, or skills of students who take an online course in their first term at the college in a systematic way. This study aimed to analyze data that the college is not currently utilizing to predict online course completion rates; furthermore, this data was used to develop an early warning system to identify students who are in danger of not completing their courses.

Introduction

Empire State College is part of the State University of New York (SUNY) system, and helps serve the state's nontraditional, adult population. The institution was founded in 1971 as a comprehensive college within the SUNY system. The college's longstanding mission is to serve adult students who require alternatives to the traditional schedule associated with higher education. The typical Empire State College student is a busy adult with a job, family responsibilities, and a schedule that does not allow for a conventional college experience. Most students study part-time and are New York State residents. The average age of an undergraduate student was 36 in the 2014-15 academic year.

There are seven things that we know about Empire State College in 2015, either from previous research or due to our students' typical characteristics:

1) The majority of the student population at the college is considered at-risk (i.e., non-traditional, adult, Pell recipients).
2) Online course completion rates are lower than the completion rates for non-online courses.
3) Course completion rates are lower for new students than for continuing students.
4) Students who do not complete all of their first-term courses are unlikely to graduate.
5) The college is not assessing the attributes, knowledge, or skills of students who take an online course in their first term at the college in a systematic way.
6) We do not know a student's education history until s/he completes the degree planning process. The majority of undergraduate students design their own individualized degree program. This process is part of a required course called Educational Planning. It is only during this degree design process that transcript credits are evaluated in order to determine what prior credit can be incorporated into the student's degree program. This degree program is submitted to the Office of College-wide Academic Review, where it must be approved. This information is recorded in the enterprise data system only after it is approved. The timing of when students take this course and submit their degree program can vary greatly, but at the earliest it is not completed until the end of a student's first term.
7) Learning Management System analytics are utilized to track faculty activity in Moodle, but not student activity within courses.

These facts led to our initial research question:

> Research Question: Can we use data that the college is not currently collecting or utilizing to predict online course completion rates; if so, can we use that information to

develop an early warning system to identify students who are in danger of not completing their courses?

Literature Review

Online Education

The Integrated Postsecondary Data System (IPEDS) estimates that over five million students (out of approximately 21 million in the higher education student body) were learning online in 2013 (Allen & Seaman, 2014, p. 14). The most recent data from the 2014 Survey of Online Learning conducted by the Babson Survey Research Group show that the number of higher education students taking at least one distance education course in 2014 was 3.7% greater than the previous year. While growth rates have slightly declined in the past six years, this growth rate is still greater than the growth rate of the overall higher education student population (Allen & Seaman, 2014, p. 14).

While the body of research on online learning in the higher education community is large (although not necessarily rigorous), there are many mixed results in terms of learning outcomes and course completion when compared to traditional classroom studies (Jaggars & Bailey, 2010, p. 1). Parkes, Stein, and Reading (2015) note that while the current generation of learners is often referred to as 'digitally native' due to their ease and familiarity with technology, the question still remains about how prepared students are for the online learning environment (p. 1). Supporters of online education argue that higher online dropout rates are due to the characteristics of students who choose online courses, rather than due to the online education itself (e.g., Howell, Laws, & Lindsay, 2004, as cited in Jaggars & Bailey, 2010).

Dray, Lowenthal, Miszkiewicz, Ruiz-Primo, and Marczynski (2011) recommend, "Given continued growth in online learning as well as reports of high attrition rates in it, understanding student readiness for online learning is necessary" (p. 29). Parkes et al. (2015) confirm that little research has been done on the preparedness or readiness of students for online learning environments. This type of information may potentially be a valuable contribution for predictive modeling of course outcomes. The ability to predict student outcomes is an important strategy that can allow instructors to identify at-risk students in order to provide timely interventions (Bienkowski, Feng & Means, 2012, as cited in Xing, Guo, Petakovic, & Goggins, 2015, p. 168).

Readiness for Online Education

The development of instruments for the assessment of online learner readiness may influence the retention and success rates of students pursuing online education (Watkins, Leigh, & Triner, 2004). One commonly used survey is McVay's Readiness for Online Learning Questionnaire (2000); it has been utilized by multiple researchers and fares well in reliability analysis (Bernard, Brauer, Abrami, & Surkes, 2004; Smith, 2005; Smith, Murphy, & Mahoney, 2003). While there are other surveys like McVay's that have been reused in several studies, universities often prefer to develop their own homegrown instruments that reflect their unique institution and online programs (Farid, 2014).

Farid's (2014) systematic review of online readiness assessment tools shows that a student's readiness for online education is a multidimensional construct that generally includes computer self-efficacy, self-direction, motivation, interaction, and attitude. Researchers have studied online readiness, defined as being ready for and open to an online learning environment (e.g., Harrell, 2008; Yukselturk, 2009); self-efficacy, defined as having confidence with necessary computer and Internet skills for online learning (e.g., Vekiri & Chronaki, 2008; Wang

& Newlin, 2002); and self-regulation, defined as having an ability for organizing and controlling behaviors, motivation and thoughts (e.g., Bol & Garner, 2011; Sun & Rueda, 2012; Yukselturk & Bulut, 2007). Research has shown that these factors can be key predictors for success in online education settings (as cited in Yukselturk & Top, 2013).

Self-direction, or self-management of learning, is a particularly predominant theme in much of the distance education literature (Calder, 2000; Evans, 2000; Warner et al., 1998; as cited in Smith et al., 2003, p. 63). Willingness to engage with others through electronic communication (participation) may also predict success in online learning (Bernard et al., 2004; Smith, 2005). Relevant research has suggested several other factors that are discriminating for predicting success in online education, including previous grade point average, study environment, age, background preparation, and access to appropriate infrastructure and associated technology (Muse, 2003; Pillay, Irving, & Jones, 2007).

At a fundamental level, learning is about how students interact and engage with subject matter, fellow classmates, and instructors. Historically, a lack of knowledge about the ways students interact with learning materials in an online environment has been one of the most significant challenges facing the field of distance education (Mattingly, Rice, & Berge, 2012, p. 238). Parkes et al. (2015) state that one factor limiting the ability of educators to determine who will be a successful distance learner is that there has been a focus placed on what students have to be (e.g., self-directed, self-aware, motivated) rather than what students need to do (p. 2). They state that this is particularly problematic because traits and characteristics are resistant to change and are difficult to develop and measure. Their suggestion for dealing with this quandary is to measure student behavior, which they hypothesize, can be adapted.

Many of these influential factors are explicit and easy to measure, such as family responsibilities and academic support. Others are more difficult to measure, like motivation. Academic motivation is positively associated with academic performance, achievement, and the "will to learn" (Singh, Singh, & Singh, 2012, p. 20). Consequently, poor motivation has been identified as an element that contributes to high dropout rates from online courses (Muilenburg & Berge, 2005).

Muilenburg and Berge (2005) evaluated both success factors and barriers, as viewed from a student perspective, which might affect learning outcomes such as learning effectiveness and motivation. Their results show that survey respondents with high levels of confidence and comfort using online learning technologies perceived significantly fewer barriers associated with online learning (p. 38). There was a significant drop in perceived barriers when a respondent had completed just one online course; moreover, as respondents reported having completed more online courses previously, their ratings of the perceived barriers decreased (p. 44).

Transcript Data

Transcripts can be important objective data sources that help track the "academic momentum of students: complex movement patterns through the curriculum that could be forward, backward, static, or all three together in any one term" (Adelman, 2006, as cited in Hagedorn & Kress, 2008, p. 8). These data can mark student engagement and answer many questions one may have about a student's prior educational history (Hagedorn & Kress, 2008).

Furthermore, studies show that academic history can predict future outcomes. Bumpus (2014) found that variables such as pre-transfer GPA and number of transfer hours predicted post-transfer outcomes for community college students (pp. 115-116). Research by List and Nadasen (2014) also found that previous GPA and credits transferred were significant predictors of retention. In addition, the number of failed courses has been shown to have a negative impact on the likelihood of graduation (Bumpus, 2014, p. 116). These types of data could contribute to the predictive power of an early warning model.

Learning Management Systems

A learning management system (LMS) is a platform designed to provide educators, administrators, and learners with a single robust, secure and integrated system to create personalized learning environments. It helps to create courses and store students' educational data on a longitudinal scale (Thakur, Olama, McNair, Sukumar, & Studham, 2014, p. 2). One way to measure student engagement in a course (i.e., how much students do) is to mine the vast amounts of data generated by interactive learning management systems, a practice that is gaining significant popularity across the higher education landscape (Buerck et al., 2013). This course management data can be used in a process called learning data analysis, or learning analytics, to search for patterns and underlying information in learning processes; the main goal is to improve learning outcomes and the learning process in online education (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014, p. 542).

Numerous studies have indicated that a positive correlation exists between student activity in an LMS and final course grades (Kotsiantis, Tselios, Filippidi, & Komis, 2013; Smith, Lange, & Huston, 2012). Macfadyen and Dawson (2010) found that the total number of discussion messages and replies posted, the total number of mail messages sent, the total number of online sessions initiated, the total number of files viewed, total time spent online, and the total number of assessments attempted and completed within a course-specific LMS were closely linked with students' final grades. Smith et al. (2012) also found that the frequency with which students logged into their LMS and how often they engaged in the material were significant factors in predicting their performance in a course. Moreover, Agudo-Peregrina et al. (2014) found that students' active participation and student-teacher interactions were among the factors most closely associated with final course grades.

Other studies indicate that the inclusion of LMS data with other sources of student data greatly enhances the ability to predict course outcomes. Campbell, Finnegan, and Collins (2006) demonstrated that adding LMS login data to students' SAT scores tripled the predictive power of their statistical model (as cited by Macfadyen & Dawson, 2010). Smith et al. (2012) found that LMS activity markers were better predictors of course outcomes than current credit load and credit completion rate in previous courses at the institution, even as early as the eighth day of the course. Other studies suggest that LMS data can be used to make predictions about at-risk students in the early stages of a course; analytics might then be used to initiate an intervention designed to change student behavior and improve learning (Mattingly et al., 2012, p. 239; Thakur et al., 2014). Ultimately, LMS data may be useful in improving student success and increasing retention (Olmos & Corrin, 2012, as cited in Dietz-Uhler & Hurn, 2013). However, even with a substantial amount of data available through learning management system usage logs, this method has been under-utilized in online education research (Agudo-Peregrina et al., 2014).

New Student Assessment

Methodology

Survey Development. In order to assess students' readiness for online learning, we developed a New Student Assessment that includes items regarding demographic information not otherwise collected by the college (e.g., marital status, number of dependents). Additionally, existing online readiness surveys and relevant literature were used for reference in the development of the questions (Bernard, Brauer, Abrami, & Surkes, 2004; California State University Stanislaus, n.d.; Central Washington University, n.d.; Dray et al., 2011; Farid, 2014; Florida Gulf Coast University, 2005; Florida Gulf Coast University, n.d.; Kerr, Rynearson, & Kerr, 2006; McVay, 2000; Miltiadou & Yu, 2000; Pillay, Irving, & Tones, 2007; Pintrich & De Groot, 1990; Smarter Measure, n.d., Smith, 2005; Smith et al., 2003; Southern Arkansas University, n.d.; University of Wisconsin Oshkosh, n.d.; University of Wisconsin Whitewater, n.d.; Watkins, Leigh, & Triner, 2004).

The final items included on the assessment fell into the following constructs: demographics, motivation, learning style, technical skills, self-efficacy, academic preparation, ability to concentrate, attitudes toward online learning, study environment, and time management skills. The instrument was tested within the institutional research office and received approval through the college's institutional review board (IRB). For the full instrument, see Appendix A.

Survey Sample and Implementation. Due to lower course completion rates for new students and online courses compared to continuing students and non-online courses, this study focused on new students enrolled in at least one undergraduate online course during the summer 2015 term (8-week term: May 18 – July 10; 15-week term: May 18 – August 28). Both matriculated and non-matriculated students were included. The final sample consisted of 400 students out of 530 new students in this term, as of May 6th. The survey was administered using the office's SurveyMonkey account. The initial invitation was sent out on May 7th, with two reminders on May 11th and May 14th. The survey closed on Friday, May 15th, prior to the start of the term.

Replacing Missing Values. There were few missing values in the New Student Assessment results. Respondents who did not complete any of the self-assessment portion of the survey were removed from the final sample used for analysis (three respondents total). Sample mode was used to replace missing values for the demographic, computer access, previous institution(s), and online course participation items. Little's MCAR test was used to determine that the missing values in the self-assessment were missing completely at random (p-value not significant). The expectation maximization estimation function in SPSS was then used to impute these missing values.

Results

There were 118 responses to the survey. The raw results are presented; missing data were imputed prior to final analysis. Results for the demographic, computer access, previous institutions, and online course participation questions are presented in Table 1. Results for the self-assessment questions are presented in Table 2.

Table 1

New Student Assessment Results – Demographics, Computer Access, Previous Institution(s), & Online Course Participation

	n *(%)*	n *(%)*	n *(%)*	n *(%)*	n *(%)*	n *(%)*	n *(%)*
Question/Category	**0**	**1-9**	**10-19**	**20-29**	**30-39**	**40+**	**Missing**
# worked/week	17 (*14.4*)	2 (*1.7*)	11 (*9.4*)	16 (*13.6*)	18 (*15.3*)	53 (*44.9*)	1 (*0.8*)
# volunteered/week	57 (*48.3*)	49 (*41.5*)	6 (*5.1*)	1 (*0.8*)	3 (*2.5*)	0 (*0.0*)	2 (*1.7*)
	0	**1**	**2**	**3+**	**Missing**		
# dependents	57 (*48.3*)	22 (*18.6*)	20 (*16.9*)	17 (*14.4*)	2 (*1.7*)		
	Middle School	**High School**	**College/ beyond**	**Other/ unknown**			
Highest level of education/parents	2 (*1.7*)	42 (*35.6*)	65 (*55.1*)	9 (*7.6*)			
	Divorced	**Married**	**Separated**	**Single**	**Widowed**		
Current marital status	11 (*9.3*)	43 (*36.4*)	3 (*2.5*)	59 (*50.0*)	2 (*1.7*)		
	Yes	**No**				**4**	
Home computer	117 (*99.2*)	1 (*0.8*)	*If No, how many times/week you can access a computer			1 (*100.0*)	
	Yes	**No**				**Yes**	**No**
*If Yes, reliable connection & speed	113 (*96.6*)	4 (*3.4*)	*If No, does this computer have reliable connection & good speed			1 (*100.0*)	0 (*0.0*)
	Yes	**No**					
Attend previous institution	97 (*82.2*)	21 (*17.8*)					
	3.5+	**3.0-3.4**	**2.5-2.9**	**2.0-2.4**	**1.5-1.9**	**<1.5**	**Missing**
*If Yes, GPA	33 (*34.0*)	32 (*33.0*)	14 (*14.4*)	10 (*10.3*)	6 (*6.2*)	0 (*0.0*)	2 (*2.1*)
	50+	**40-49**	**30-39**	**20-29**	**11-19**	**1-9**	**0/Missing**
*If Yes, # credits	63 (*64.9*)	6 (*6.2*)	11 (*11.3*)	7 (*7.2*)	5 (*5.2*)	2 (*2.1*)	3 (*3.1*)
	<1 yr.	**1-2 yrs.**	**2-3 yrs.**	**3-4 yrs.**	**4-5 yrs.**	**5+ yrs.**	
*If Yes, time since last college course	39 (*33.1*)	12 (*10.2*)	11 (*9.3*)	3 (*2.5*)	3 (*2.5*)	29 (*24.6*)	
	No	**Yes, college**	**Yes, at job**	**Yes, college & job**	**Yes, other**		
Previous participation in online course	46 (*39.0*)	41 (*34.7*)	4 (*3.4*)	17 (*14.4*)	10 (*8.5*)		
	1	**2**	**3-4**	**5+**	**Missing**		
"Yes, college": # college courses	9 (*22.0*)	6 (*14.6*)	14 (*34.1*)	11 (*26.8*)	1 (*2.4*)		
	90-100 (A)	**80-89 (B)**	**70-79 (C)**	**60-69 (D)**	**Less than 60 (F)**		
"Yes, college": College course GPA	19 (*46.3*)	20 (*48.8*)	2 (*4.9*)	0 (*0.0*)	0 (*0.0*)		
	1	**2**	**3-4**	**5+**			
"Yes, job": # courses for job	1 (*25.0*)	0 (*0.0*)	1 (*25.0*)	2 (*50.0*)			
	1	**2**	**3-4**	**5+**			
"Yes, college & job": # college courses	1 (*5.9*)	5 (*29.4*)	4 (*23.5*)	7 (*41.2*)			
	90-100 (A)	**80-89 (B)**	**70-79 (C)**	**60-69 (D)**	**Less than 60 (F)**		
"Yes, college & job": College course GPA	10 (*58.8*)	5 (*29.4*)	2 (*11.8*)	0 (*0.0*)	0 (*0.0*)		
	1	**2**	**3-4**	**5+**			
"Yes, college & job": # courses for job	2 (*11.8*)	4 (*23.5*)	1 (*5.9*)	10 (*58.8*)			
	1	**2**	**3-4**	**5+**			
"Yes, other": # courses	4 (*40.0*)	2 (*20.0*)	3 (*30.0*)	1 (*10.0*)			

Table 2

New Student Assessment Results – Self-assessment

Item	Strongly Agree n (%)	Agree n (%)	Slightly Agree n (%)	Slightly Disagree n (%)	Disagree n (%)	Strongly Disagree n (%)	Missing n
I am confident that I can pay for my education at Empire State College.	47 (*40.9*)	39 (*33.9*)	16 (*13.9*)	6 (*5.2*)	5 (*4.3*)	2 (*1.7*)	3
I am self-motivated.	65 (*56.5*)	41 (*35.7*)	5 (*4.3*)	2 (*1.7*)	0 (*0.0*)	2 (*1.7*)	3
I am comfortable learning new technologies.	63 (*54.8*)	44 (*38.3*)	5 (*4.3*)	0 (*0.0*)	0 (*0.0*)	3 (*2.6*)	3
I am comfortable participating in an online discussion.	48 (*41.7*)	51 (*44.3*)	12 (*10.4*)	0 (*0.0*)	1 (*0.9*)	3 (*2.6*)	3
I am comfortable working with computers.	66 (*57.4*)	37 (*32.2*)	9 (*7.8*)	1 (*0.9*)	1 (*0.9*)	1 (*0.9*)	3
I am confident in my ability to excel in an online course.	56 (*48.7*)	42 (*36.5*)	9 (*7.8*)	6 (*5.2*)	0 (*0.0*)	2 (*1.7*)	3
I am confident that I can do college-level work.	66 (*57.9*)	44 (*38.6*)	3 (*2.6*)	0 (*0.0*)	0 (*0.0*)	1 (*0.9*)	4
I am confident that I will complete my courses this term.	72 (*62.6*)	37 (*32.2*)	5 (*4.3*)	0 (*0.0*)	0 (*0.0*)	1 (*0.9*)	3
I am effective in communicating my opinion in writing to others.	55 (*47.8*)	47 (*40.9*)	8 (*7.0*)	3 (*2.6*)	0 (*0.0*)	2 (*1.7*)	3
I am good at completing tasks independently.	68 (*59.6*)	40 (*35.1*)	2 (*1.8*)	1 (*0.9*)	1 (*0.9*)	2 (*1.8*)	4
I believe that my background and experience will be beneficial to my studies.	71 (*61.7*)	35 (*30.4*)	6 (*5.2*)	0 (*0.0*)	1 (*0.8*)	2 (*1.7*)	3
I can complete my work even when there are distractions.	42 (*36.8*)	50 (*43.9*)	15 (*13.2*)	2 (*1.8*)	3 (*2.6*)	2 (*1.8*)	4
I can stay focused on a task when necessary.	50 (*43.9*)	57 (*50.0*)	4 (*3.5*)	1 (*0.9*)	0 (*0.0*)	2 (*1.8*)	4
I feel that online learning is of equal quality or higher quality than traditional classroom learning.	36 (*31.9*)	41 (*36.3*)	21 (*18.6*)	9 (*8.0*)	4 (*3.5*)	2 (*1.8*)	5
I have significant experience using a Learning Management System (Moodle, etc.)	34 (*29.6*)	26 (*22.6*)	27 (*23.5*)	7 (*6.1*)	8 (*7.0*)	13 (*11.3*)	3
I have enough time to study for my course(s).	34 (*29.6*)	55 (*47.8*)	19 (*16.5*)	3 (*2.6*)	3 (*2.6*)	1 (*0.9*)	3
I have good time management skills.	40 (*34.8*)	50 (*43.5*)	18 (*15.7*)	2 (*1.7*)	3 (*2.6*)	2 (*1.7*)	3
I have the technical skills necessary to complete online courses.	63 (*54.8*)	40 (*34.8*)	8 (*7.0*)	2 (*1.7*)	0 (*0.0*)	2 (*1.7*)	3
I typically complete assignments on time.	65 (*57.0*)	43 (*37.7*)	4 (*3.5*)	1 (*0.9*)	0 (*0.0*)	1 (*0.9*)	4
I usually study in a place where I can concentrate on my coursework.	51 (*44.7*)	49 (*43.0*)	9 (*7.9*)	3 (*2.6*)	1 (*0.9*)	1 (*0.9*)	4
My main goal this term is gaining a thorough understanding of the material that will be covered in my course(s).	64 (*56.1*)	44 (*38.6*)	3 (*2.6*)	0 (*0.0*)	1 (*0.9*)	2 (*1.8*)	4
My main goal this term is getting good grades.	67 (*58.3*)	41 (*35.7*)	3 (*2.6*)	2 (*1.7*)	1 (*0.9*)	1 (*0.9*)	3

Transcript Data Collection

Methodology

Transcript data were recorded for survey respondents who also were matriculated students during the summer 2015 term. Course information was entered into an online form on SurveyMonkey at the registration level (see Appendix B). This information included course subject, course level, credits, grade received, and OPE ID. An OPE ID is an identification number used by the U.S. Department of Education to identify schools that have a Program Participation Agreement, which allows its students to be eligible to participate in Federal Student Financial Assistance programs under Title IV regulations (OPE ID, n.d.).

Various categories were created using this aggregated transcript data. Certain categories are related to a student's Area of Study (AOS). An AOS at Empire State College is the equivalent of a major, in a broader sense. For example, there is a Science, Mathematics and Technology (SMT) AOS. This is more comprehensive than a major would be; a student in this AOS would have a more specific concentration, such as biology or computer science. A cross-walk was created that matched the subject of the transcript course with the AOS that the student entered when they matriculated at Empire State College (e.g., courses designated as math, natural sciences, or applied sciences were matched with the SMT AOS).

The categories include:

- Credits attempted, completed, and eligible for transfer (at Empire State College, this is a grade of C or higher)
- Percent of courses incomplete; percent of courses failed
- Credit and course completion rates

- Overall GPA out of 4.0
- Time since last institution (end of last term at previous institution estimated using the end date for the equivalent term at Empire State College; start date of summer 2015 term)
- Credits attempted and completed in **last year** (isolated data from student's last previous term, and any activity in the two immediately preceding terms (e.g., if a student's last term was spring 2015, these data were aggregated with fall 2014 and summer 2014 registrations, if present)
- Credit and course completion rates in **last year**
- Percent of courses that matched **AOS/major**
- **AOS/major** course completion rate and **AOS/major** GPA out of 4.0
- Percent of math courses that were failed or incomplete
- Percent of writing courses that were failed or incomplete

Results

There were 87 matriculated students out of the 118 survey respondents (73.7%), and 67 of these students (77.0%) had transcript information recorded in our Nolij database. There were 78 different institutions represented: 22 SUNY community colleges; 10 SUNY four-year schools, 10 City University of New York (CUNY) schools, and 36 other institutions (9 of which were for-profits). Previous institution information is presented in Table 3. Aggregated transcript data are presented in Table 4.

Table 3

Previous Institutions Attended

	1	2	3	4	5
Category	n (*%*)	n (*%*)	n (*%*)	n (*%*)	n (*%*)
# of overall previous institutions attended per student	38 (*56.7*)	18 (*26.9*)	7 (*10.4*)	3 (*4.5*)	1 (*1.5*)

	SUNY CC	**SUNY 4-yr.**	**CUNY**	**Other**	**(For-profit subset)**
	n (*%*)	n (*%*)	n (*%*)	n (*%*)	n (*%*)
Proportion of sample who attended	35 (*52.2*)	12 (*17.9*)	14 (*20.9*)	29 (*43.3*)	10 (*14.9*)*

* The "for-profit" column (*n* = 10, 14.9%) is a subset of the "Other" category (i.e., 10 of the 29 students who attended an institution in the "Other" category attended a for-profit. A total of 19 of these students did not attend a for-profit.).

Table 4

Aggregated Transcript Data (n = 67)

Category	**Mean**	**SD**	**Minimum**	**Maximum**
Credits attempted	88.8	43.6	9.0	212.0
Credits completed	73.5	35.4	9.0	167.0
Credits transferable (C & up)	65.4	32.1	9.0	149.0
Courses incomplete, *%*	*11.5%*	*13.6%*	*0.0%*	*58.8%*
Courses failed, *%*	*7.9%*	*10.1%*	*0.0%*	*40.0%*
Credit Completion Rate	*84.5%*	*15.7%*	*28.3%*	*100.0%*
Course Completion Rate	*80.6%*	*17.6%*	*29.4%*	*100.0%*
Overall GPA (out of 4.0)	2.74	0.70	1.57	4.00
Time since last institution transcript	7.2 yrs.	8.9 yrs.	0.0 yrs.	34.0 yrs.
Last Year, Credits attempted	22.4	13.2	0.0	66.0
Last Year, Credits completed	18.7	13.5	0.0	66.0
Last Year, Credit Completion Rate	*79.6%*	*28.3%*	*0.0%*	*100.0%*
Last Year, Course Completion Rate	*78.2%*	*27.9%*	*0.0%*	*100.0%*
AOS/major Course Match Rate	*27.1%*	*27.5%*	*0.0%*	*100.0%*
AOS/major Course Completion Rate	*76.6%*	*30.8%*	*0.0%*	*100.0%*
AOS/major GPA (out of 4.0)	2.64	1.08	0.00	4.00
Math courses, % failed or incomplete	*30.8%*	*31.5%*	*0.0%*	*100.0%*
Writing courses, % failed or incomplete	*19.0%*	*29.0%*	*0.0%*	*100.0%*

Learning Management System Data

Methodology

A total of 118 students completed the New Student Assessment in May 2015. One respondent did not have a valid email tied to his/her response and therefore could not be used in the analysis. Respondents who did not complete any of the self-assessment portion of the survey were also removed from the final sample used for analysis (three respondents in total). As a result, four responses were removed from the file, resulting in 114 students. Of those 114 students, 103 students were still enrolled in an undergraduate online course at the conclusion of add/drop week for the summer 2015 term. These students took a total of 197 undergraduate online courses during this term. Of these 197 registrations, 145 resulted in "credit," which was defined as a passing letter grade or a grade of "full-credit," and 52 registrations resulted in "no credit," which was defined as a grade of "no credit," "incomplete," or "withdrawal;" a course completion rate of 73.6%.

ESC utilizes Moodle as its learning management system to deliver online undergraduate courses. In this initial assessment of LMS data, we focused our efforts on whether or not a student logged into their course, made a discussion post, or viewed a discussion. We excluded the time prior to the start of the course and week 1 (add/drop week) because of the high volume of registration activity within undergraduate online courses (i.e., students moving into and out of course sections). We excluded weeks 9-15 from this analysis because the summer term at the college provides students with an option to take 8- and 15-week courses. The majority of undergraduate online courses at the college consist of modules or sections that students must complete to receive "credit" for taking the course. However, this course design does not preclude students from working ahead and finishing the final course module before the final week of the course. As a result, week 8 was excluded from this analysis as well.

Results

The percentage of registrations made by students who were active within their course through the LMS decreased from week 2 through week 7. Complete results are presented in Table 5.

Table 5
Course Activity within LMS, Weeks 2-7

	Logins	**Posts**	**Views**
Course Week	n *(%)*	n *(%)*	n *(%)*
Week 2	194 *(98.5)*	145 *(73.6)*	189 *(95.9)*
Week 3	177 *(89.8)*	112 *(56.9)*	171 *(86.8)*
Week 4	173 *(87.8)*	128 *(65.0)*	164 *(83.2)*
Week 5	164 *(83.2)*	97 *(49.2)*	150 *(76.1)*
Week 6	164 *(83.2)*	112 *(56.9)*	155 *(78.7)*
Week 7	159 *(80.7)*	100 *(50.8)*	146 *(74.1)*

Because of our interest in using LMS data to create an early warning system, we wanted to focus more closely on students' activity within their courses in weeks 2 and 3 of the course. Nearly 90% of registrations were made by students who logged into their course through the LMS in both weeks, while a slightly lower percentage were made by students who viewed a discussion in both weeks. Less than one-half of registrations were made by students who made a post in both weeks 2 and 3 of the course. Complete results are presented in Table 6.

Table 6
Course Activity within LMS, Weeks 2-3

	Logins	**Posts**	**Views**
	n *(%)*	n *(%)*	n *(%)*
0 weeks	3 *(1.5)*	29 *(14.7)*	7 *(3.6)*
1 week	17 *(8.6)*	79 *(40.1)*	20 *(10.2)*
2 weeks	177 *(89.8)*	89 *(45.2)*	170 *(86.3)*

Statistical Modeling

Methodology

Variable Selection. Within our sample of 197 registrations, we identified 42 categorical variables, which allowed us to observe a statistically significant difference on course completion rates between groups. Variables were coded so the reference group (largest group) appeared last among the categories. More descriptive variable names were used in these tables than in previous tables to provide additional context. These results are depicted in Tables 7a-7d.

Table 7a
Variables from College's Database where Statistically Significant Differences Existed in Course Completion Rates between Groups

Variable	Variable Categories	n (%)	ASR	$\chi 2$	V
First term enrollment status	Part-time	82 *(84.1)*	2.8	8.04*	0.20
	Full-time	115 *(66.1)*	-2.8		
Subject	Science/Math/Tech	44 *(56.8)*	-2.9	8.72**	0.21
	Business	29 *(82.8)*	1.2		
	Human Services	21 *(81.0)*	0.8		
	Arts and Humanities	103 (76.7)	1.0		

Note. ** = $p < 0.01$, * = $p < 0.05$. ASR=Adjusted standardized residuals.

Table 7b

Variables from the New Student Assessment where Statistically Significant Differences Existed in Course Completion Rates between Groups

Variable	Variable Categories	n (%)	ASR	χ2	V
50+ transfer credits	No previous	28 *(92.9)*	2.5	6.34*	0.18
	< 50	64 *(71.9)*	-0.4		
	50+	105 *(69.5)*	-1.4		
GPA from previous institutions	Not reported/no previous	34 (91.2)	2.6	26.53***	0.37
	1.5-2.9	51 (47.1)	-5.0		
	3.0 +	112 (80.4)	2.5		
I am comfortable participating in online discussions.	Strongly Agree (6)	92 *(81.5)*	2.4	5.57*	0.17
	Strongly Disagree to Agree (1-5)	105 *(66.7)*	-2.4		
Institution before ESC	No	28 *(92.9)*	2.5	6.23*	0.18
	Yes	169 *(70.4)*	-2.5		
I feel that online learning is of equal quality or higher quality than traditional classroom learning.	Strongly disagree to slightly agree (1-4)	67 *(62.7)*	-2.5	6.23*	0.18
	Agree to strongly agree (5-6)	130 *(79.2)*	2.5		
I have good time management skills.	Strongly disagree to slightly agree (1-4)	47 *(59.6)*	-2.5	6.25*	0.18
	Agree to strongly agree (5-6)	150 *(78.0)*	2.5		
Marital Status	Divorced, Married, Separated, Widowed	98 *(81.6)*	2.5	6.47*	0.18
	Single	99 *(65.7)*	-2.5		
Number of online courses previously taken	No previous	69 *(78.3)*	1.1	13.41**	0.26
	1-2 courses	44 *(52.3)*	-3.6		
	3 or more courses	84 *(81.0)*	2.0		
Time since last institution	No previous	28 *(92.9)*	2.5	15.33***	0.28
	More than 2 years	79 *(59.5)*	-3.7		
	2 years or fewer	90 *(80.0)*	1.9		

Note. *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$. ASR=Adjusted standardized residuals

Table 7c

Variables from the Transcript Data where Statistically Significant Differences Existed in Course Completion Rates between Groups

Variable	Variable Categories	n (%)	ASR	χ2	V
AOS/major attempted credits from prior institution(s)	No AOS credits	36 *(77.8)*	0.6	23.11**	0.34
	Above Median	42 *(47.6)*	-4.3		
	Below Median	49 *(71.4)*	-0.4		
	No transcript credits	70 *(88.6)*	3.5		
AOS/major completed credits from prior institution(s)	No AOS credits	36 *(77.8)*	0.6	20.12**	0.32
	Above Median	43 *(51.2)*	-3.8		
	Below Median	48 *(68.8)*	-0.9		
	No transcript credits	70 *(88.6)*	3.5		
AOS/major course completion rate from prior institution(s)	No AOS credits	36 *(77.8)*	0.6	31.95**	0.40
	Above Median	41 *(80.5)*	1.1		
	Below Median	50 *(44.0)*	-5.5		
	No transcript credits	70 *(88.6)*	3.5		
AOS/major course completion rate from prior institution(s)	No AOS credits	36 *(77.8)*	0.6	31.95**	0.40
	Above Median	41 *(80.5)*	1.1		
	Below Median	50 *(44.0)*	-5.5		
	No transcript credits	70 *(88.6)*	3.5		
AOS/major GPA from prior institution(s)	No AOS credits	36 *(77.8)*	0.6	44.03**	0.47
	Above Median	43 *(86.0)*	2.1		
	Below Median	48 *(37.5)*	-6.5		
	No transcript credits	70 *(88.6)*	3.5		
Attempted credits in year prior to transferring to ESC	No transcript	70 *(88.6)*	3.5	16.00*	0.28
	Below Median	56 *(57.1)*	-3.3		
	Above Median	71 *(71.8)*	-0.4		
Attended a City University of New York institution prior to ESC	No transcript credits	70 *(88.6)*	3.5	16.20**	0.29
	Yes	34 *(52.9)*	-3.0		
	No	93 *(69.9)*	-1.1		
Attended a for-profit institution prior to ESC	No transcript credits	70 *(88.6)*	3.5	12.95*	0.26
	Yes	17 *(58.8)*	-1.4		
	No	110 *(66.4)*	-2.6		
Attended a non-State University of New York or a City University of New York institution prior to ESC	No transcript credits	70 *(88.6)*	3.5	17.67*	0.30
	Yes	56 *(55.4)*	-3.7		
	No	71 *(73.2)*	-0.1		
Attended a State University of New York community college prior to ESC	No	61 *(52.5)*	-4.5	22.57**	0.34
	Yes	66 *(77.3)*	0.8		
	No transcript	70 *(88.6)*	3.5		
Attended a State University of New York four-year college prior to ESC	No transcript	70 *(88.6)*	3.5	12.78*	0.25
	Yes	23 *(69.6)*	-0.5		
	No	104 *(64.4)*	-3.1		
Credit completion rate in last year prior to transferring to ESC	Above Median	58 *(79.3)*	1.2	23.22**	0.34
	Below Median	69 *(53.6)*	-4.7		
	No transcript	70 *(88.6)*	3.5		
Credit completion rate from prior institution(s)	No transcript	70 *(88.6)*	3.5	29.21**	0.39
	Above Median	55 *(83.6)*	2.0		
	Below Median	72 *(51.4)*	-5.4		
Course completion rate in last year prior to transferring to ESC	No transcript	70 *(88.6)*	3.5	22.44**	0.34
	Above Median	51 *(80.4)*	1.3		
	Below Median	76 *(55.3)*	-4.6		
Credits attempted at prior institution(s)	Below Median	59 *(72.9)*	-0.2	15.73**	0.28
	Above Median	68 *(58.8)*	-3.4		
	No transcript	70 *(88.6)*	3.5		
Credits completed at prior institution(s)	Below Median	58 *(70.7)*	-0.6	14.08**	0.27
	Above Median	69 *(60.9)*	-3.0		
	No transcript	70 *(88.6)*	3.5		
Credits completed in last year	Below Median	57 *(54.4)*	-3.9	18.92**	0.31

prior to transferring to ESC	Above Median	70 *(74.3)*	0.2		
	No transcript	70 *(88.6)*	3.5		
Number of previous institutions attended	No transcript	70 *(88.6)*	3.5	17.20*	0.30
	2 institutions	29 *(55.2)*	-2.4		
	3 institutions	19 *(63.2)*	-1.1		
	4 institutions	4 *(100.0)*	1.2		
	5 institutions	2 *(50.0)*	-0.8		
	1 institution	73 *(68.5)*	-1.2		
Overall course completion rate from prior institution(s)	No transcript	70 *(88.6)*	3.5	27.13**	0.37
	Above Median	53 *(83.0)*	1.8		
	Below Median	74 *(52.7)*	-5.2		
Overall GPA from prior institution(s)	Above Median	62 *(80.6)*	1.5	27.10**	0.37
	Below Median	65 *(50.8)*	-5.1		
	No transcript credits	70 *(88.6)*	3.5		
Percentage of failed courses from prior institution(s)	Below Median	58 *(81.0)*	1.5	26.03**	0.36
	Above Median	69 *(52.2)*	-5.0		
	No transcript	70 *(88.6)*	3.5		
Percentage of incompletes from prior institution(s)	Below Median	50 *(78.0)*	0.8	19.31**	0.31
	No transcript	70 *(88.6)*	3.5		
	Above Median	77 *(57.1)*	-4.2		
Percentage of math courses failed/incomplete from prior institution(s)	No Math courses	17 *(64.7)*	-0.9	12.64*	0.25
	Above Median	47 *(63.8)*	-1.7		
	Below Median	63 *(66.7)*	-1.5		
	No transcript credits	70 *(88.6)*	3.5		
Percentage of total credits at prior institutions within AOS/major	No transcript credits	70 *(88.6)*	3.5	13.15*	0.26
	Above Median	55 *(61.8)*	-2.3		
	Below median	72 *(68.1)*	-1.3		
Percentage of writing courses failed/incomplete from prior institution(s)	No Writing courses	7 *(100.0)*	1.6	17.92**	0.30
	Above Median	46 *(58.7)*	-2.6		
	No transcript credits	70 *(88.6)*	3.5		
	Below median	74 *(66.2)*	-1.8		
Time since last institution	No transcript	70 *(88.6)*	3.5	16.00**	0.28
	Above Median	56 *(57.1)*	-3.3		
	Below Median	71 *(71.8)*	-0.4		
Transcript submitted	No	70 *(88.6)*	3.5	12.52**	0.25
	Yes	127 *(65.4)*	-3.5		
Transferable credits from prior institution(s)	Below Median	59 *(67.8)*	-1.2	12.86*	0.26
	Above Median	68 *(63.2)*	-2.4		
	No transcript	70 *(88.6)*	3.5		

Note. ** = $p < 0.001$, * = $p < 0.01$. ASR=Adjusted standardized residuals

Table 7d

Variables from LMS Data where Statistically Significant Differences Existed in Course Completion Rates between Groups

Variable	Variable Categories	n (%)	ASR	$\chi 2$	V
Number of weeks logging into course in weeks 2 and 3	0 weeks	3 *(0.0)*	-2.9	33.65*	0.41
	1 week	17 *(23.5)*	-4.9		
	2 weeks	177 *(79.7)*	5.7		
Number of weeks making a discussion post in weeks 2 and 3	0 weeks	29 *(24.1)*	-6.5	48.67*	0.50
	1 week	79 *(73.4)*	0.0		
	2 weeks	89 *(89.9)*	4.7		
Number of weeks viewing a discussion in weeks 2 and 3	0 weeks	7 *(14.3)*	-3.6	42.82*	0.47
	1 week	20 *(25.0)*	-5.2		
	2 weeks	170 *(81.8)*	6.5		

Note. * = $p < 0.001$. ASR=Adjusted standardized residuals

Within these 42 variables there were numerous variables that were redundant with one another. One example of this redundancy is: 1) we asked students on the survey to estimate their GPA from their previous institution(s) and 2) we computed an overall GPA for students based on their transcript data. Another example: 1) we computed credit completion rates from transcript data and 2) we computed course completion rates from transcript data. A third example of redundancy involved looking at the number of weeks within weeks 2 and 3 of the course when a student: 1) logged into the course, 2) made a discussion post, and 3) viewed a discussion. The reason using all three variables as predictors is problematic is that to view a discussion, a student must log into their course and to make a post, a student must view a discussion.

As a result, we selected the variable among the redundant group of variables that produced the largest difference in course completion rates across groups according to the effect size measure (Cramer's V) as a result of a Pearson chi square test. This allowed us to eliminate a total of 16 variables; leaving us with a total of 26 variables. This process is depicted in Table 8. Our reasons for doing this were twofold: 1) we wanted to minimize the impact of collinearity within our model, and 2) we wanted to maintain an appropriate cases to variables ratio. The rule of thumb for logistic regression is a minimum of 10 outcome events per predictor variable (Vittinghoff & McCulloch, 2006).

Table 8
Variables Retained and Eliminated from the Dataset

Retained	**Eliminated**
Flag denoting whether a student had transcript data *(transcript)*	1) Flag denoting whether a student attended an institution prior to ESC *(survey)*
Number of weeks within weeks two and three that a student made a discussion post *(LMS)*	1) Number of weeks within weeks 2 and 3 that a student logged into their course *(LMS)* and 2) Number of weeks within weeks 2 and 3 that a student viewed a discussion *(LMS)*
Credits attempted *(transcript)*	1) Credits completed *(transcript),* 2) Transferable credits *(transcript),* and 3) 50 plus credit flag *(survey)*
Time since last institution *(transcript)*	1) Time since last institution *(survey)*
Credits attempted within AOS/major *(transcript)*	1) Credits completed within AOS/major *(transcript)* and 2) Percent of total credits taken within AOS/major *(transcript)*
Credit completion rate *(transcript)*	1) GPA *(transcript),* 2) Course completion rate *(transcript),* and 3) GPA from previous institutions *(survey)*
AOS/major GPA *(transcript)*	1) AOS/major credit completion rate *(transcript)* and *2)* AOS/major course completion rate *(transcript)*
Credits completed within last year *(transcript)*	1) Credits attempted within last year *(transcript)*
Credit completion rate within last year *(transcript)*	1) Course completion rate within last year *(transcript)*

Note: The data source for each variable is in parenthesis and italicized.

Modeling. To begin creating our statistical model, we selected only those registrations from our sample (n=197), which resulted in “no credit” (n=52). Then, we took a random sample of 52 registrations from the 145 registrations that resulted in “credit” and merged the two files together. This resulted in a file with 104 registrations; 52 registrations that resulted in “credit” and 52 registrations that resulted in “no credit.” We then created three more files following the same methodology. This left us with four 50/50 training datasets.

Our next step was to run a binary logistic regression in SPSS on all four training datasets using a flag denoting whether or not the registration resulted in “credit” as the dependent variable and all 26 of the aforementioned variables as covariates. We selected “forward conditional” as our method to ensure that we would keep a favorable cases to variables ratio. Each training dataset produced between four and eight models. With regard to selecting the most accurate model from the training datasets, we gave preference to models that were more accurate in predicting the outcome of the registrations that resulted in “no credit.” In other words, we were willing to sacrifice some overall accuracy to gain greater accuracy in predicting that a student would not complete his or her course.

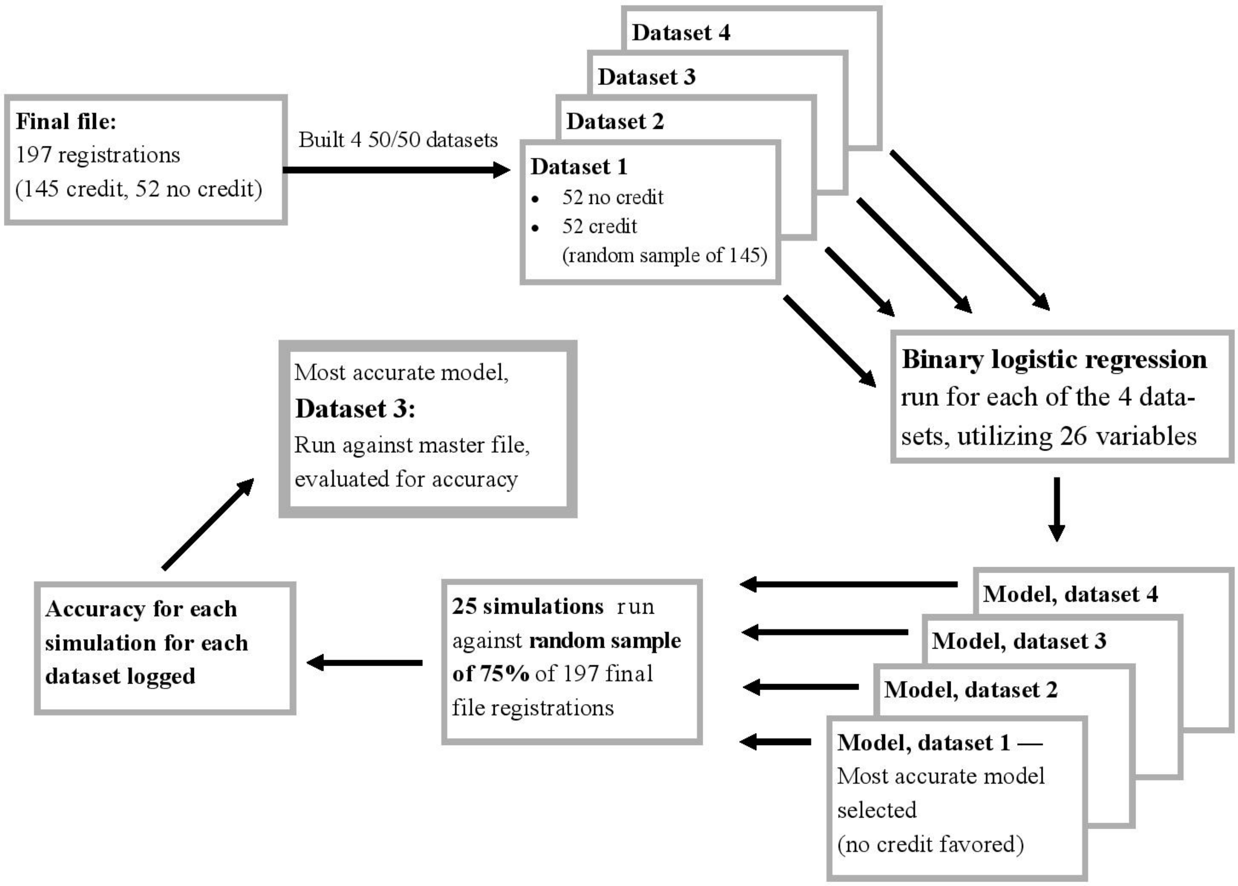
The models produced by training datasets two and three consisted of three variables; the model produced by the first training dataset consisted of four variables; and the model produced by the third training dataset consisted of eight variables. The variables *number of weeks making a post in weeks 2 and 3* and *AOS/major GPA at previous institutions* were present in all four models, while the variables *number of previous online courses (as assessed by the survey)* and whether or not a student *attended a State University of New York four-year institution* were present in two of the four models. Complete results for each model are depicted in Table 9.

Table 9
Models Produced by Training Datasets

Training Dataset	Variables
1	1) Number of weeks making a discussion post in weeks 2 and 3, 2) AOS/major GPA from previous institutions, 3) Number of previous institutions, and 4) Number of previous online courses
2	1) Number of weeks making a discussion post in weeks 2 and 3, 2) AOS/major GPA from previous institutions, and 3) SUNY four-year attendance flag
3	1) Subject, 2) First-term enrollment status, 3) Number of weeks making a discussion post in weeks 2 and 3, 4) AOS/major GPA from previous institutions, 5) Percentage of incompletes from previous institutions, 6) SUNY four-year attendance flag, 7) Comfort level participating in online discussions (survey), and 8) Marital status (survey)
4	1) Number of weeks making a discussion post in weeks 2 and 3, 2) AOS/major GPA from previous institutions, and 3) Number of previous online courses (survey)

We then ran 25 simulations for each of the four models produced by our training datasets on a random sample of 75% of our overall sample of 197 registrations (i.e., natural distribution). Again, we selected a binary logistic regression in SPSS using the variable denoting whether or not the registration resulted in "credit" as the dependent variable and then only the variables from the models produced by each of our training datasets as covariates. In this instance, rather than selecting forward conditional, we selected "enter" as our method to ensure that all of the variables from those models would be present in our simulations. We then logged the performance of each model across the 25 simulations. The entire process is depicted in Figure 1.

Figure 1



The model produced by the third training dataset was the most accurate model. Its overall accuracy was nearly 90%; however, it was much less accurate predicting the outcomes of the registrations that resulted in "no credit" than "credit." Complete results for each of the models across the 25 simulations are depicted in Table 10.

Table 10

Results of Model Testing

Training Dataset	No Credit	Credit	Overall
1	67.8%	93.6%	86.7%
2	64.7%	92.3%	84.7%
3	70.4%	95.9%	89.2%
4	62.4%	93.3%	85.2%

Our final step in the modeling process was to test this model on our sample of 197 undergraduate online registrations made during the summer 2015 term. Again, we selected a binary logistic regression in SPSS using the variable denoting whether or not the registration resulted in "credit" as the dependent variable and all of the variables from the model produced by the third training dataset as covariates. Again, we selected "enter" as our method to ensure that all of the variables from this model would be present.

Results

The model accurately predicted the outcome of 174 of 197 registrations (88.3%). The model accurately predicted 35 of the 52 registrations (67.3%) that resulted in "no credit" and 139 of the 145 registrations (95.9%) that resulted in "credit." In total, we gained an additional 14.7 percentage points of predictive accuracy (73.6% (constant/course completion rate) → 88.3%).

Discussion and Limitations

These data indicate that we can predict course outcomes with some degree of accuracy using data that the college is not currently collecting or utilizing. That said, this is an extremely small sample size and we would want to replicate this study on a larger scale before bringing anything forward regarding the creation of an early warning system designed to predict course completion rates for new students taking undergraduate online courses. Another limitation to this study is that the population consists of students who were new to the college in the summer 2015 term. The majority of new students at ESC start in either the fall or spring terms. As a result, we may be dealing with a population that is somewhat atypical of new students at the college.

In addition, there was very little variance for the items on the New Student Assessment. This may have something to do with the fact that we administered the survey, which is in large part assessing students' technical skills and comfort level with technology, online. The very fact that the students responded to the survey tells us that they have, at the very least, basic computer skills. It also stands to reason that those students with below average technical skills who may also lack comfort with technology either chose not to respond to the survey or were unaware of our request to participate altogether. An idea that may correct this problem is to pilot a similar instrument during an onsite orientation to capture a broader array of students based on their technical abilities.

A student's educational history appears to be a strong predictor of future success. Students who had higher credit/course completion rates and grade point averages overall, within their AOS/major, and during the last year prior to starting at ESC were more likely to complete their first-term undergraduate online courses. That said, the process of manually entering these data is problematic. In addition to being extremely time consuming, there is a lack of standardization across institutions regarding transcripts. For example, colleges categorize subjects differently and utilize different coding schemes to identify course levels and term lengths. In addition, some colleges report credits, while others report credit hours. The fact that we had three staff members from our office entering data and making judgement calls as a result of this lack of standardization almost certainly compromised our reliability.

In addition, students who complete the non-matriculated student application process at ESC typically do not submit college transcripts. A total of 70 registrations from our sample were made by students who were either non-matriculated, did not attend an institution prior to attending ESC, or simply chose not to submit a transcript from a previous institution as part of the matriculated application process. The completion rate for courses in academic year 2014-15 at ESC was 81.0%. Across the college, the course completion rate for non-matriculated students was 84.1%, while the course completion rate for matriculated students was 80.8%. The overall course completion rate for our sample for the summer 2015 term was 73.6%, while the course completion rate for students with no transcript data was 88.6%. The fact that this population of students performed so well certainly increased the number of variables where statistically significant differences on course completion rates were present and most likely impacted our overall results.

Data gathered from the college's LMS tracking whether or not students were active within their course in the early weeks of the course appear to be a good predictor of course completion rates. The strongest indicator of success was whether or not a student made a discussion post within a given week.

Currently, the college is tracking the LMS activity of adjunct faculty within courses. These reports were designed by staff from Information Technology Services and included student activity as well. We extracted the student information directly from these reports. Unfortunately, the data behind these reports did not allow us to distinguish between multiple instances of the same action (i.e., login, post, view) within the same minute for the same student in a particular course. As a result, we only felt comfortable indicating whether or not a student was active within the week, rather than being able to quantify or qualify that activity. In the future our goal is to do both. In terms of quantifying actions, our plan is to standardize activity by course section by creating percentile groups of activity (i.e., 0-33%, 34-66%, 67-100%) and then tracking course completion rates by those groups. In addition, we would like to qualify discussion posts and further categorize those posts based on length and lexical diversity.

References

Agudo-Peregrina, A. F., Iglesias-Pradas, S., Conde-González, M. A., & Hernández-García, A. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*, 542-550.

Allen, I. E., & Seaman, J. (2014). Grade level: Tracking online education in the United States. Retrieved from http://onlinelearningconsortium.org/read/survey-reports-2014/

Bernard, R. M., Brauer, A., Abrami, P. C., & Surkes, M. (2004). The development of a questionnaire for predicting online learning achievement. *Distance Education*, *25*(1), 31-47.

Bol, L., & Garner, J. K. (2011). Challenges in supporting self-regulation in distance education environments. *Journal of Computing in Higher Education*, *23*(2-3), 104-123.

Buerck, J. P., Mudigonda, S. P., Mooshegian, S. E., Collins, K., Grimm, N., Bonney, K., & Kombrink, H. (2013). Predicting non-traditional student learning outcomes using data analytics-a pilot research study. *Journal of Computing Sciences in Colleges*, *28*(5), 260-265.

Bumpus, J. G. (2014). *Influences on community college transfer student persistence at an urban public university: developing and validating a predictive model using student demographic data and transcript data* (Doctoral dissertation, The University of Alabama TUSCALOOSA).

California State University Stanislaus. (n.d.). *Online readiness self-assessment.* Retrieved from http://teachonline.csustan.edu/selfassessment.php

Central Washington University. (n.d.). *Online readiness.* Retrieved from http://www.cwu.edu/online-learning/online-readiness

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*(1). Retrieved from www.ncolr.org/jiol

Dray, B. J., Lowenthal, P. R., Miszkiewicz, M. J., Ruiz-Primo, M. A., & Marczynski, K. (2011). Developing an instrument to assess student readiness for online learning: A validation study. *Distance Education*, *32*(1), 29-47.

Farid, A. (2014). Student Online Readiness Assessment Tools: A Systematic Review Approach. *The Electronic Journal of e-Learning, 12*(4), 375-382.

Florida Gulf Coast University. (2005). *Distance learning self-readiness assessment.* Retrieved from http://www.fgcu.edu/support/technologyskills/begin.asp?surveyid=3

Florida Gulf Coast University. (n.d.). *Technology skills self-assessment.* Retrieved from https://survey.fgcu.edu/Survey.aspx?s=311d40c08e234f9181d7f97e6623fbcc#

Hagedorn, L. S., & Kress, A. M. (2008). Using transcripts in analyses: Directions and opportunities. *New Directions for Community Colleges*, *2008*(143), 7-17.

Harrell, I. L. (2008). Increasing the Success of Online Students. *Inquiry*, *13*(1), 36-44.

Jaggars, S., & Bailey, T. R. (2010). Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis.

Kerr, M. S., Rynearson, K., & Kerr, M. C. (2006). Student characteristics for online learning success. *The Internet and Higher Education*, *9*(2), 91-105.

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. *International Journal of Technology Enhanced Learning*, *5*(2), 133-150.

List, A., & Nadasen, D. (2014). *Community College Transfer Students' Persistence at University.* Paper presented at 41st NEAIR Conference, Philadelphia, PA.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588-599.

Mattingly, K.D., Rice, M.C., & Berge, Z.L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal, 4*(3), 236-247.

McVay, M. (2000). *Developing a web-based distance student orientation to enhance student success in an online bachelor's degree completion program* (Doctoral dissertation, Nova Southeastern University).

Miltiadou, M., & Yu, C. H. (2000). Validation of the Online Technologies Self-Efficacy Scale (OTSES).

Muilenburg, L. Y., & Berge, Z. L. (2005). Student barriers to online learning: A factor analytic study. *Distance Education*, *26*(1), 29-48.

Muse Jr., H. (2003). The Web-based community college students: An examination of factors that lead to success and risk. *Internet and Higher Education, 6*, 241-261.

OPE ID. (n.d.). In *Integration Postsecondary Education Data System Glossary.* Retrieved November 5, 2015, from https://nces.ed.gov/ipeds/glossary/index.asp?id=433

Parkes, M., Stein, S., & Reading, C. (2015). Student preparedness for university e-learning environments. *Internet and Higher Education*, *25*, 1-10.

Pillay, H., Irving, K., & Tones, M. (2007). Validation of the diagnostic tool for assessing tertiary students' readiness for online learning. *High Education Research & Development*, *26*(2), 217-234.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Singh, S., Singh, A., & Singh, K. (2012). Motivation levels among traditional and open learning undergraduate students in India. *The International Review of Research in Open and Distributed Learning*, *13*(3), 19-40.

Southern Arkansas University. (n.d.). *Online readiness self-assessment.* Retrieved from https://web.saumag.edu/online/assessment/

Smarter Measure. (n.d.) *Readiness for education at a distance indicator.* Retrieved from http://www.readi.info/

Smith, P. J. (2005). Learning preferences and readiness for online learning. *Educational Psychology*, *25*(1), 3-12.

Smith, P. J., Murphy, K. L., & Mahoney, S. E. (2003). Towards identifying factors underlying readiness for online learning: An exploratory study. *Distance Education*, *24*(1), 57-67.

Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, *16*(3), 51-61.

Sun, J. C. Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191-204.

Thakur, G. S., Olama, M. M., McNair, A. W., Sukumar, S. R., & Studham, S. (2014). *Towards Adaptive Educational Assessments: Predicting Student Performance using Temporal Stability and Data Analytics in Learning Management Systems*. Oak Ridge National Laboratory (ORNL). Retrieved from http://aspiringminds.com/pages/assess/2014/camera_ready/poster/thakur_etal.pdf

University of Wisconsin Oshkosh. (n.d.). *Online readiness assessment.* Retrieved from http://www.uwosh.edu/llce/online-readiness-assessment-survey-1

University of Wisconsin Whitewater. (n.d.). *Instructional, communication & information technology: Getting started.* Retrieved from http://www.uww.edu/icit/olr/getting-started

Vekiri, I., & Chronaki, A. (2008). Gender issues in technology use: Perceived social support, computer self-efficacy and value beliefs, and computer use beyond school. *Computers & Education*, *51*(3), 1392-1404.

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American journal of epidemiology*, *165*(6), 710-718.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168-181.

Wang, A. Y., & Newlin, M. H. (2002). Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class. *Computers in Human Behavior*, *18*(2), 151-163.

Watkins, R. Leigh, D. & Triner, D. (2004). Assessing readiness for e-learning. *Performance Improvement Quarterly, 17*(4), 66-79.

Yukselturk, E. (2009). Do entry characteristics of online learners affect their satisfaction? *International Journal on E-Learning*, *8*(2), 263-281.

Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Journal of Educational Technology & Society*, *10*(2), 71-83.

Yukselturk, E., & Top, E. (2013). Exploring the link among entry characteristics, participation behaviors and course outcomes of online learners: An examination of learner profile using cluster analysis. *British Journal of Educational Technology*, *44*(5), 716-728.

Appendix A

SUNY Empire State College New Student Self-Assessment

Demographics

1. Number of hours you currently work in a typical week:
    a. 0
    b. 1-9
    c. 10-19
    d. 20-29
    e. 30-39
    f. 40+
2. Number of hours you spend volunteering or doing non-compensated charitable work in a typical week:
    a. 0
    b. 1-9
    c. 10-19
    d. 20-29
    e. 30-39
    f. 40+
3. Number of dependents
    a. 0
    b. 1
    c. 2
    d. 3+

4. What is the highest level of education attained by either of your parents?
    a. Middle school
    b. High school
    c. College or beyond
    d. Other/unknown
5. What is your current marital status?
    a. Divorced
    b. Married
    c. Separated
    d. Single
    e. Widowed

Computer Access

6. *Do you have a home computer
    a. No
    b. Yes
7. *If "Yes," Do you have a reliable internet connection with good speed?
    a. No
    b. Yes
8. *If "No," How many times per week can you access a computer?
    a. Fewer than once per week
    b. 1
    c. 2
    d. 3

e. 4

f. 5

g. 6

h. 7

9. *If "No," Does this computer have a reliable internet connection with good speed?

a. No

b. Yes

Previous Institution(s)

10. *Did you attend an institution before enrolling at ESC?

a. No

b. Yes

11. *If "Yes," Approximate GPA from previous institution(s)

a. Less than 1.5

b. 1.5-1.9

c. 2.0-2.4

d. 2.5-2.9

e. 3.0-3.4

f. 3.5+

12. *If "Yes," Number of total credits from previous institution(s):

a. 0

b. 1-9

c. 11-19

d. 20-29

e. 30-39

f. 40-49

g. 50+

13. *If "Yes," Time since last college course:

a. Less than 1 year

b. 1-2 years

c. 2-3 years

d. 3-4 years

e. 4-5 years

f. 5+ years

Online Course Participation

14. Have you previously participated in an online course?

a. No

b. Yes, in college

c. Yes, at my job

d. Yes, in college and at my job

e. Yes, other

15. *If "b," Number of previous online college courses:

a. 1

b. 2

c. 3-4

d. 5+

16. *If "b," Average grade in previous online college courses:

a. 90-100 (A)
b. 80-89 (B)
c. 70-79 (C)
d. 60-69 (D)
e. Less than 60 (F)

17. *If “c,” Number of online courses:
    a. 1
    b. 2
    c. 3-4
    d. 5+
18. *If “d,” Number of previous online college courses:
    a. 1
    b. 2
    c. 3-4
    d. 5+
19. *If “d,” Number of previous online courses you took for your job:
    a. 1
    b. 2
    c. 3-4
    d. 5+
20. *If “d,” Average grade in previous online college courses:
    a. 90-100 (A)
    b. 80-89 (B)

c. 70-79 (C)

d. 60-69 (D)

e. Less than 60 (F)

21. *If “e,” Number of online courses:

a. 1

b. 2

c. 3-4

d. 5+

Self-assessment

22. Please rate your level of agreement with the following statements (6 – Strongly Agree, 5 – Agree, 4 – Slightly Agree, 3 – Slightly Disagree, 2 – Disagree, 1 – Strongly Disagree)

a. I am confident that I can pay for my education Empire State College.

b. I am self-motivated.

c. I am comfortable learning new technologies.

d. I am comfortable participating in an online discussion.

e. I am comfortable working with computers.

f. I am confident in my ability to excel in an online course.

g. I am confident that I can do college-level work.

h. I am confident that I will complete my courses this term.

i. I am effective in communicating my opinion in writing to others.

j. I am good at completing tasks independently.

k. I believe that my background and experience will be beneficial to my studies.

l. I can complete my work even when there are distractions.

m. I can stay focused on a task when necessary.

n. I feel that online learning is of equal quality or higher quality than traditional classroom learning.

o. I have significant experience using a Learning Management System (Moodle, Angel, Blackboard, Desire2Learn, Edmodo, myCourses, etc.).

p. I have enough time to study for my course(s).

q. I have good time management skills.

r. I have the technical skills necessary to complete online courses.

s. I typically complete assignments on time.

t. I usually study in a place where I can concentrate on my coursework.

u. My main goal this term is gaining a thorough understanding of the material that will be covered in my course(s).

v. My main goal this term is getting good grades.

Appendix B

New Student Assessment Database

1. Student ID: ____________
2. Institution ID: ____________
3. Term
    a. Fall
    b. Spring
    c. Summer
4. Year
    a. 1950

       …

    b. 2015
5. Subject Area
    a. Business
    b. Arts & Humanities
    c. Math
    d. Education
    e. Health Sciences & Medicine
    f. Physical/Health Education
    g. Natural Sciences
    h. Applied Sciences & Technology (e.g., engineering, computer science, architecture)
    i. Social & Behavioral Sciences

j. Writing & Reading

k. Trades & Technical Skills

l. Miscellaneous

6. Course level

a. 100

b. 200

c. 300

d. 400

e. 500

f. 600

g. 700

h. 800

7. Credits

a. 0

b. 0.5

c. 1

d. 1.5

e. 2

f. 2.5

g. 3

h. 4

i. 4.5

j. 5

k. 6

l. 7

m. 8

n. 9+

8. Grade

a. A

b. B

c. C

d. D

e. E

f. F

g. Satisfactory/pass

h. Unsatisfactory/fail

i. Did not complete (e.g., withdrawal)

Leading Institutional Change from Below:

A Case Study

Carol Van Zile-Tamsen

State University of New York at Buffalo

Presented at the Annual Meeting of the

Northeast Association of Institutional Research

November 2, 2015

Abstract

In July, 2014, a large research institution in the northeast created the Center for Educational Innovation. The Center was the culmination of a three-year effort to change the campus culture such that faculty members and administrators would come to value teaching and assessment of student learning a more integral part of the university mission. Two committees, one task force, and countless hours of advocacy eventually achieved the desired result. The purpose of this paper is to illustrate how avenues for shared leadership, along with advocacy efforts, can be leveraged to bring about meaningful institutional change, even within a highly regulated public institution. This manuscript concludes with generalized strategies for bringing about change on campuses, regardless of campus culture, sector, or institutional control.

Leading Institutional Change from Below:

A Case Study

In July, 2014, the Center for Educational Innovation was created at a large public research institution in the northeastern United States. This Center serves as serves as a nexus for campus-wide efforts to further elevate the scholarship of, and research support for, pedagogical advancement and improved learning at the university. Center staff members are committed to advancing the scholarship of teaching and learning through integrated services, education, research and development related to university teaching, learning, and assessment. The Center was the culmination of a three-year effort to change the campus culture such that faculty members and administrators would come to value teaching and assessment of student learning as a more integral part of the university mission. Since the institution is a member of the Association of American Universities (AAU) and a high research institution, the importance of planful pedagogy and assessment of student learning often seemed to get lost in the push to advance the research mission.

Change in institutions of higher education can move at a very slow pace for several reasons. First, these institutions are large bureaucracies, with many policies and procedures in place (Scott, 1998). In addition, the traditional shared governance model, where significant change must be approved by key stakeholders, is still valued and still used in everyday decision making (Caruth & Caruth, 2013). The purpose of this manuscript is to illustrate how these same avenues for shared leadership, along with individual change agents willing to invest in advocacy efforts, can be leveraged to bring about meaningful and necessary institutional change, even within a highly regulated and bureaucratic public institution. This process for bringing about change at this institution is described within theoretical frameworks of institutional change, addressing external factors, as well as internal culture, policies, and politics. The institution is the unit of focus, and the theoretical concepts are illustrated with examples from the unique reality and culture of this institution, addressing a very real gap in existing research in organizational change in higher education (Fumasoli & Stensaker, 2013).

Barriers to Change in Higher Education

As mentioned above, institutions of higher education are often large, bureaucratic institutions steeped in tradition. For these reasons, meaningful change can be difficult to achieve. However, additional organizational features also contribute to resistance to change, and, in fact, often serve as barriers to meaningful change. In this section, several barriers that often prevent cultural changes in higher education are outlined.

Institutions of higher education, especially large ones, are often organized vertically into silos (Keeling, Underhile, & Wall, 2007). The purpose for organizing into silos was to encourage disciplines to act with autonomy and creativity in a way that was separate from other disciplines. That goal was definitely achieved as academic units organized along disciplinary lines often view themselves as autonomous organizations, acting independently from other academic units and even from central administration. Formal communications often circulate vertically through the siloes, with the most value being given to intra-silo communication. It is much more difficult to circulate communications horizontally, especially when inter-silo communications are often given less weight than communications that originate within the academic unit.

At most colleges and universities, the shared governance model is still a valued tradition (Caruth & Caruth, 2013). Including faculty in decision making through a faculty governance body and through the use of committees to develop and review policy is a common practice.

One of the most basic problems of this approach to institutional governance is simply that it takes a lot of time. However, in many cases, faculty serving in governance roles or sitting on committees often do not have content expertise related to the areas where they are making decisions, which is why a great deal of time is often spent doing background research. There is often an implied goal to maximize consensus and minimize conflict. In the end, the decision making process has very little to do with efficiency, and often the groups involved can lose sight of what is best for the institution as they get caught up in the political balancing act of the needs of the stakeholders involved.

Within this system, institutions have faculty members who have complete job security (Martensson, 2015). Those who have achieved tenure are likely to stay at an institution through retirement and often see the many changes being pushed by central administration as temporary, whimsical initiatives that will soon blow over as soon as the next big issue comes along (Henard & Roseveare, 2012). For that reason, it is very difficult to get them enthused to change anything that does not have a direct impact on their own interests. In some systems, professional tenure (i.e., permanent appointment) is also possible for administrative staff. In such a situation, faculty and staff alike can be unmotivated to embrace new initiatives.

The barriers to change that are engendered by the system of faculty tenure are often exacerbated by the faculty reward system (Martensson, 2015). Faculty are primarily rewarded for research productivity through their publications and research funding, and this is becoming increasingly true for institutions that are not categorized as research institutions (The Teagle Working Group on the Teaching Scholar, 2007). Activities related to teaching, assessment of student learning, and service to the institution that are outside the bounds of activities that get rewarded are often ignored and devalued. Efforts to change the culture to focus on these activities in addition to research are futile unless underlying reward systems are also changed (Ginsberg & Bernstein, 2011; Martensson, 2015).

Adding to the difficulty of motivating faculty to engage in activities that are typically outside of the faculty reward system and viewed as fleeting whims of central administration is the status of the academy as a not-for-profit, mission-driven organization. In this setting, faculty members are expected to work independently and creatively. They often see themselves apart from the more mundane administrative tasks of the institution (Caruth & Caruth, 2013; Henard & Roseveare, 2012). While this environment has contributed to a focus on vision and values rather than the bottom line, leadership strategies to encourage faculty members to act in accordance with institutional vision and values must be more cooperative and less directive than are found in for-profit companies.

This mission-driven focus, in conjunction with a vertical structure, a shared governance model, a tenure system and a faculty reward system that recognizes faculty research productivity above all else presents a context in which change can almost never happen in a quick or meaningful fashion without drastic measures. Further, many faculty members have also come to use the mantra of academic freedom as a shield to resist or ignore mandates that originate from central administration (Caruth & Caruth, 2013; Henard & Roseveare, 2012). This is particularly true if the mandates are not linked into the faculty reward system.

To further complicate matters, there is the constant conflict between the shared institutional history, which includes past failed efforts to enact change, and the current institutional change trajectory. Roxa and Martensson (2011) characterize these two institutional aspects as the Saga versus the Enterprise. These authors suggest that an overreliance on the saga and on formal avenues of authority to mandate change that can interfere with the institution's ability to embark upon and achieve its enterprise.

One additional way to think about barriers to institutional change in higher education is to think about frames of organizational change (Bolman & Deal, 2008). These include: human resources (adding staff and/or training them appropriately), structural (assigning appropriate roles and responsibilities), political, and symbolic (culture and values). When the focus is placed on one of these over others, change can be difficult, at best. Instead, all of these areas should be addressed when a meaningful and long-lasting change in culture is trying to be achieved.

Leveraging Strengths to Overcome Saga and Achieve the Enterprise

As the previous section has outlined, there are many barriers to meaningful and long-lasting change in higher education (Bolman & Deal, 2008; Caruth & Caruth, 2013; Ginsberg & Bernstein, 2011; Henard & Roseveare, 2012; Keeling et al., 2007; Martensson, 2015; Roxa & Martensson, 2011; The Teagle Working Group on the Teaching Scholar, 2007), and the points outlined in the preceding section are likely not the components of an exhaustive list. However, there are examples of successful change from which to draw illustrative principles for overcoming these barriers and achieving meaningful and long-lasting institutional change.

Andrade (2011) suggests using the multiple frames outlined by Bolman and Deal (2008) to manage change. In her paper, she specifically addresses the creation of a culture of assessment in which faculty develop a greater commitment to assessing student learning. For example, in terms of getting faculty more involved in assessment work, she suggests that a human resources approach might involve providing training for faculty in various aspects of assessment while also creating a larger vision for the importance of faculty participating in assessment work so that they understand what a crucial role they play in ensuring that students are learning. In terms of the symbolic frame, she suggests integrating assessment work into the vision and values of the institution via the institutional mission statement, as well as developing a culture of assessment by instituting award ceremonies, transition rituals, and showcase events.

Ginsberg and Bernstein (2011) describe three roles that are essential when trying to bring about culture change on a campus. These include Leaders, Change Agents, and Facilitators. The Leaders possess institutional power and/or authority to help change culture, while Facilitators have an institutional role that gives them some measure of authority combined with some measure of expertise. Change Agents don't have any formal power or authority. They are "on-the-ground" experts who walk-the-walk and have passion to lead the culture shift. It is essential to engage individuals in all three of these roles in any attempt to change campus culture.

Either formal identification of change agents or their self-identification is crucial for enacting change on a campus. In an article describing how the Bologna Process was used to reform higher education in Italy, Ballarino and Perotti (2012) conclude by suggesting that any analysis of change in higher education should include an attempt to characterize the actors involved, along with their behaviors and interactions, since these are absolutely essential to the process. However, individual attempts of individual and disconnected change agents working in isolation on grass-roots efforts will likely not have sufficient momentum to make a strong impact (Roxa & Martensson, 2011).

In the Hannah and Lester (2009) framework of organizational learning, these individual change agents would be seen as the micro-level of the organization. The administrators at the central level of the institution and at the highest levels of leadership within each academic level are operating at the macro-levels of organizational learning. Mandated change comes from these levels, and for all the reasons described in the preceding section, is often ignored or treated as a reporting requirement that will go away when the next mandate comes along (Henard & Roseveare, 2012; Martensson, 2015).

Grass roots level change comes from the lowest levels of the institution, individual faculty and staff members at the micro level (Hannah & Lester, 2009). This type of change can happen but is often disorganized and disconnected – maybe too disconnected to have large impacts (Martensson, 2015). The meso level of institutional learning contains all of the mid-level leaders and all of the interconnections that these leaders have with individuals at the micro and the macro levels of the institution. Martensson and Roxa (2015) suggest that meso level changes (inter connected networks with macro and micro level partners) seem to offer the best hope of true culture change. Ginsberg and Bernstein (2011) would concur since this is the level where their change facilitators "live."

It appears that the value of the meso level approach to culture change is two-fold. First, the individuals who "live" at this level of the institution have a legitimate authority to lead and make decisions. Having such legitimacy is important in encouraging followers to invest time and energy in an initiative. They will be more likely to do so if they feel the initiative has a chance of being viewed as important by institutional leadership. Second, these individuals also tend to have content area knowledge and can be considered experts. This places them firmly in the facilitator role (Ginsberg & Bernstein, 2011). They can then begin to bring the grass roots leaders and the administrative leadership together in a top-down and a bottom-up fashion simultaneously to influence culture from both sides at once. Further, they can use these meso-networks to negotiate horizontally across silos, using collaborative microcultures to communicate in a way that top-down administrative mandates never can (Henard & Roseveare, 2012).

Individuals leading change in meso-networks and those at the microlevel who are identified or self-identified to be facilitators and change agents must possess key characteristics to be effective, however. First, they need to be politically astute and aware of the institutional history. Second, they need to have charisma and enthusiasm. Third, they need to have good communication and advocacy skills. Finally, they need to balance tact with assertiveness. In sum, with dynamic and persuasive individuals leading change efforts through the use of focused meso-networks that cut across and through traditional organizational structures, meaningful and long-lasting change can happen.

A Case Study: Efforts to Enact Long-Lasting and Meaningful Change

In this section of the paper, an actual case of meaningful and, hopefully, long-lasting change will be described, showing how these principles were brought to bear to create the Center for Educational Innovation and begin to change the culture at a large research intensive institution. It is important to note, however, that the actors in this example were doing their best with the circumstances they found themselves in to do right by the institution. At no time, did they scour the literature to find out what worked elsewhere or what should work in theory to bring about change and apply it their situation. All of the application of change concepts from the literature has been done after the fact.

In the present case, the shared governance model provided an opportunity for a group of like-minded faculty and staff from a variety of departments, all members of a standing committee devoted to promoting a culture of assessment and institutional improvement, to creatively address two major gaps in resources and support for faculty: pedagogical assistance and support for conducting assessment work at the level required for regional accreditation. Several of these individuals were meso-level leaders with connections to micro and macro level leaders. Several of these individuals were micro-level change agents, experts in pedagogy and/or assessment with charisma and enthusiasm and willing to lead a grass level movement to positively impact the student experience at the institution.

These committee members were able to garner support by taking advantage of external stressors and internal organizational turnover. Further, individual committee members used advocacy skills to lobby for the desired results. The creation of this new unit was the end result of a three-year effort that occurred during a time of significant institutional activity and transformation that included: (a) ascendance of the sitting provost to office of the president, (b) the arrival of a new provost from outside the institution, (c) reorganization of the Office of the Provost, (d) preparation for regional reaccreditation, and (e) re-envisioning of the strategic plan.

Understanding the Institutional Saga

At the beginning of this effort, the institutional focus was very much on research and economic development and still is. This institution is a member of the American Association of Universities and is considered a high research institution. In 2003, the arrival of a new president marked a strategic planning process in which the focus was to grow the research enterprise, increase institutional efficiency so as to invest savings into strategic initiatives, and on the growth of the knowledge economy as a driver for regional economic growth and rebirth. Academic excellence remained a key component of the university mission statement, but the efforts were all focused on the research enterprise.

Throughout the strategic planning process and the implementation of the strategic plan, two important dichotomies came to the forefront that led to confusion and mixed messages with regard to what the trues values were and what activities would be rewarded. First, an intensive process was carried out to identify Strategic Strength Areas, which would serve as key areas of focus for interdisciplinary research and lay the foundation to attract large federal grant awards. At the same time, a large group of faculty and professional staff were engaged in an effort to propose programs and activities that would lead to an excellent education and an enhanced student experience. Since the underlying faculty reward system was never altered to include rewards relating to teaching and mentoring students, the focus for most faculty members remained on research productivity.

Second, there was an emphasis on working together horizontally, across academic units, to form Strategic Strength Areas, but in reality the formal communication channels remained vertical and the independent silos remained the primary source of allegiance. The concept of "One University" was slow to gain traction; for those interdisciplinary researchers who received funding, the allegiance was to each other and to the research and the Strategic Strength Area rather than to central administration.

Within this institutional environment, where the leadership was focused on research and economic development and believing that the educational mission would take care of itself, regional accreditation reporting with increased accountability for student learning outcomes was no easy task. The mid-point accreditation report was due during this time frame, and the institution received recommendations with regard to both the goals and assessment of its general education program and the overall assessment of student learning outcomes. Soon after the recommendations from the regional accreditor were received, the sitting president announced his retirement, and the sitting provost was named the new president. The institution began organizing to respond to the accreditation report and prepare for its decennial review, and the assessment steering committee was formed. It was at this point that the meso-leaders and change agents, who serendipitously formed the committee, realized that they had an opportunity to enact needed change at the institution, and they made a plan to act.

The Unit and Faculty Context

The meso-leaders consisted of the associate director of assessment, the associate dean of undergraduate education, and the associate dean of graduate education. The change agents included faculty from across the institution. In their early meetings, what they immediately reported was that the academic units openly operated as independent units, often trying to ignore central mandates. Any efforts by the central administration to "force" units to participate in general education assessment or be more proactive in student learning assessment would likely fail. Related to this fact, many faculty members seemed to mistrust central services and supports, particularly when the words "assessment, evaluation, and documentation" were used. Even if academic unit leaders were willing to help push forward central assessment and improvement initiatives, faculty were not willing to get involved for fear of negative consequences. Further, the tenure and promotion system and the faculty reward system would not allow them to be recognized for any quality efforts in these areas, so there were no advantages. For junior faculty, especially, the mantra was to "worry about tenure then worry about teaching."

Change Strategies

<u>Shared Governance.</u> In truth, the shared governance approach that set this change effort in motion was part of a strategic effort to engage middle-level leaders, who held key roles and varying levels of assessment expertise, with faculty, also with various levels of assessment expertise. These individuals were members of a single assessment steering committee but also served on various working teams for the accreditation self-study. The group further included key

representatives from the Faculty and Professional Staff Senates. In the end, the assessment steering group consisted of nearly 40 members with representatives from every academic unit and from central administration. The membership included experts in program assessment, accreditation reporting, traditional pedagogy, and online learning. There were also several representatives from student service areas, such as from the libraries and student affairs.

Advocacy. The wide reach of the committee meant that advocacy and outreach efforts by individuals could play a very significant role in changing the culture and promoting the idea of a dedicated center for pedagogy and assessment. Individuals used their knowledge of accreditation requirements and the gaps identified in the self-study to demonstrate for leaders that many faculty members needed support to become better teachers and to understand and conduct assessment of student learning at the level necessary to meet accreditation requirements. The change agents and meso-leaders also understood that most faculty members needed to hear top-down messages that teaching and assessment are important activities that *will* be rewarded; thus, they helped senior leaders draft memos and web sites and even outlined potential low-cost rewards and recognitions programs. At the same time, the same committee members worked with other faculty members, as well as staff in academic units, to help them understand the importance of ensuring student learning, the importance of accreditation to the institution, and, the role of every employee in helping achieve a successful outcome for every student. The idea was to focus on the importance of teaching quality for students and trying to keep the focus off the needs of the central administration.

The key middle managers were strong advocates for a culture of assessment and putting the focus back on teaching and learning, and the faculty change agents were in complete agreement. As a result, in addition to all of the individual and small group meetings that were taking place, the middle managers organized trainings for faculty in the area of assessment since that was the area where faculty seemed to need the greatest amount of understanding. These town halls were basically large introductory classes on the assessment cycle. Then, these were followed up a month later with smaller workshops on more focused areas of assessment, including the use of rubrics and curriculum mapping, to help faculty get into the details of how they would actually conduct assessment activities in a class. Following these workshops, smaller, on-demand sessions were conducted for individual academic units, departments, programs, and individual faculty as needed. Additional materials were posted online. These sessions were followed up with sessions on how to make changes to courses and programs based on assessment results.

Once most programs were well under way with regard to annual program assessment and review for improvement, and the self-study working teams had finalized their overall report on the status of assessment for the reaccreditation self-study, the assessment steering committee stopped to take stock of the status of teaching, learning, and assessment of student learning at the institution. At this point, the committee members had spent over 18 months working to support the needs of the campus with regard to assessment and improving teaching to improve learning. All members but one (the associate director of assessment) were doing this work as institutional service, and it was becoming apparent that the institution needed a cadre of professionals who were paid to support faculty 100% of the time in their pedagogy and assessment efforts. It was at this point that the committee developed written recommendations to merge the existing teaching and learning center, which focused primarily on classroom technology support, with the office of assessment, into a comprehensive center of pedagogy and assessment.

Capitalizing on external forces. The written report was completed and submitted to a key vice provost within one year of the decennial reaccreditation visit. At that point, it was still unclear if all of the efforts after the mid-term report recommendations had been received would be sufficient to help the institution achieve full accreditation from the regional accreditor. Thus, the written recommendations of the assessment steering committee included clear demonstrations of how the creation of a merged and enhanced center of pedagogy and assessment would support the institution in its reaccreditation efforts.

Results of Efforts to Enact Change

The advocacy effort to the vice provost was successful. He read and digested the initial report and recommendations and saw the value of such a center. He then gathered a subcommittee of the assessment steering committee to further refine the recommendations and present a recommendation for a new center for pedagogy and assessment to the provost. The provost then decided to include pedagogy and assessment in the re-envisioning of the strategic plan. Rather than reconstituting a new committee to explore the feasibility of a unit to support faculty in these areas, they turned to the same committee, asking the members to refine their recommendations for the re-envisioning of the strategic plan.

In the end, the entire senior leadership team supported the proposal to merge the existing teaching and learning center and assessment office into a new and much improved office to support pedagogy and assessment. When the accreditation team came for the decennial review, they recommended that not only should the institution support this center but should fully fund it, as well. The institution followed those recommendations, and the new center has been in place for sixteen months and has been making great strides in changing the culture by engaging faculty in the scholarship of teaching and learning (Ginsberg & Bernstein, 2011). Not only is there an effort to have assessment viewed as just one aspect of good teaching, but also as research in which the faculty member collects data about the effects of various instructional strategies on learning outcomes. There is also an effort to help faculty see that teaching excellence is related to innovation.

Conclusions

There are many barriers to change in institutions of higher education that result from the way they are structured to the way that faculty are rewarded to the way that decisions are made (Caruth & Caruth, 2013; Keeling et al., 2007; The Teagle Working Group on the Teaching Scholar, 2007). From the outside, based in some cases on the lack of quality decision-making, and in others on the length of the decision-making process or on the types of decisions that are made, it can appear that the organization and leadership in higher education is "organized anarchy" (Fumasoli & Stensaker, 2013). However, it is important to understand that higher education cannot be understood in the same way that businesses and for-profit companies are understood. Mandated change is often ignored or treated as a reporting requirement that will go away when the next mandate comes, while grass roots level change can happen but is often too disorganized and disconnected to have institutional impacts (Martensson, 2015). Meso level changes -- inter connected networks of middle level managers leading macro and micro level change agents -- seem to offer the best hope of true culture change (Ginsberg & Bernstein, 2011; Martensson, 2015).

In the illustrated case of institutional change presented here, the shared governance model was effective to begin the process of culture change at this large research-intensive campus because it was composed of exactly the types of networks that both Martensson (2015) and Ginsberg and Bernstein (2011) describe. The individuals who were chosen as members of the assessment steering committee were either in the appropriate positions of authority or had appropriate expertise, or both, to lead efforts to raise awareness about the need for greater focus on pedagogy and assessment at the institution. A core group of them were committed to improving the campus culture in these areas and were charismatic and enthusiastic leaders, willing to engage with faculty and staff members, as well as senior leaders, to promote a culture of improved teaching and assessment.

Their efforts were convincing to senior leaders, who in turn, purposefully used this committee and its expertise to address continuing questions related to pedagogical support and assessment of student learning. As a result, the wheel wasn't reinvented each time the institution needed to examine this issue during the period of self-study review for accreditation and strategic plan re-envisioning. The respect for the committee's work was evidenced by the support its recommendations for a center of pedagogy and assessment received from senior leadership and from the accreditation review team.

While the institution still has some work to do in terms of fully engaging faculty in the scholarship of teaching and learning, there is a much stronger culture of assessment with the creation of the center and a much stronger interest in improving teaching and innovating pedagogy. The new center is fully staffed, with a director, an associate director, 14 full-time staff members, one part-time staff member, a full-time graduate assistant, a part-time graduate assistant, and three work-study students, and demand continues to grow.

The provost awarded the center $50,000 per year for three years to begin a seed grant program to fund small grants in the area of teaching and learning. In the first year of the program, ten seed grant projects were funded for $5,000 each to research the impacts of various innovative instructional techniques. The center collaborates with student affairs to sponsor Assessment Day, a professional development day for faculty and staff members devoted to all areas of assessment. Registration has grown across the three years of the event, with the most recent event having over 220 registrants. The annual seminar on teaching excellence sponsored by the center most recently had 90 registrants. Based on these initial indicators, it seems that faculty members are beginning to think more about their teaching and about student learning outcomes and how to assess them. Based on the positive results achieved at this large, research-intensive institution, the same method of using multilevel networks with key meso-leaders is recommended for other types of institutions as a way to initiate meaningful and long-lasting change.

References

Andrade, M. S. (2011). Managing Change -- Engaging Faculty in Assessment Opportunities. *Innovative Higher Education, 36*, 217-233.

Ballarino, G., & Perotti, L. (2012). The Bologna Process in Italy. *European Journal of Education, 47*(3), 348-363.

Bolman, L. G., & Deal, T. E. (2008). *Reframing Organizations: Artistry, Choice, and Leadership*. San Francisco, CA: Wiley.

Caruth, G. D., & Caruth, D. L. (2013). Understanding Resistance to Change: A Challenge to Universities. *Turkish Online Journal of Distance Education, 14*(2), 12-21.

Fumasoli, T., & Stensaker, B. (2013). organizational Studies in Higher Education: A Reflection on Historical Themes and Prospective Trends. *Highed Education Policy, 26*, 479-496.

Ginsberg, S. M., & Bernstein, J. L. (2011). Growing the Scholarship of Teaching and Learning Through Institutional Culture Change. *Journal of Scholarship of Teaching and Learning, 11*(1), 1-12.

Hannah, S. T., & Lester, P. B. (2009). A Multilevel Approach to Building and Leading Learning Organizations. *The Leadership Quarterly, 20*, 34-48.

Henard, F., & Roseveare, D. (2012). Fostering Quality Teaching in Higher Education: Policies and Practices. Paris, France.

Keeling, R. P., Underhile, R., & Wall, A., F. (2007). Horizontal and Vertical Structures. *Liberal Education, 93*(4), np.

Martensson, K. (2015). *Local Level Leadership -- Crucial for SOTL?* Paper presented at the International Society for the Scholarship of Teaching and Learning, Melbourne, Australia.

Martensson, K., & Roxa, T. (2015). Leadership at a Local Level -- Enhancing Educational Development. *Educational Management Administration & Leadership, NV*, 1-16. doi: DOI: 10.1177/1741143214549977

Roxa, T., & Martensson, K. (2011). Understanding Strong Academic Microcultures. Lund, Sweden: Lund University.

Scott, W. R. (1998). *Organizations: Rational, Natural, and Open Systems*. Upper Saddle River, NJ: Prentice-Hall.

The Teagle Working Group on the Teaching Scholar. (2007). Student Learning and Faculty Research: Connecting Teaching and Scholarhship. New York, NY: American Council of Learned Societies.

Increasing Connections to Increase Online Student Retention

Heidi M. Baez

City University of New York School of Professional Studies

Evgeniya Reshetnyak

Fordham University

Jennifer Sparrow

City University of New York School of Professional Studies

Author Note

Heidi M. Baez, Institutional Research, City University of New York School of Professional Studies

Evgeniya Reshetnyak, Graduate Student, Fordham University

Jennifer Sparrow, Department of General Education, City University of New York School of Professional Studies

This research was supported by the City University of New York Office of Academic Affairs Student Success Grant.

Correspondence concerning this article should be addressed to Heidi M. Baez, Institutional Research, CUNY School of Professional Studies, 119 West 31st Street, New York, NY 10001

Contact: Heidi.Baez@mail.cuny.edu

Abstract

With an increase in the population of degree completers accessing education online, institutions must thoughtfully address the needs of this population, starting at the entry point. At the CUNY School of Professional Studies we hypothesized that online students' persistence and performance for the first term could be improved by implementing structured activities during the new student orientation that build connections to peers, faculty and profession. This newly designed orientation foregrounds development of interpersonal and disciplinary connections designed to provide students with access to social support, information, and resources that will assist in developing their professional identities. Interactions via the orientation site Discussion Boards were coded for quality and quantity to operationalize the connections new students were making during the orientation period. Using a combination of logistic regression (for persistence outcomes) and ANCOVA and multiple regression models (for academic performance outcomes) analyses, we found that both increased interactions amongst students, their peers and facilitators; and connections to the discipline had positive relationships with persistence and performance outcomes. Students who made more interpersonal connections had higher first-term GPAs and re-enrolled at a higher rate. Additionally, students who were able to clearly communicate professional outlooks had higher GPA outcomes as well. This study also confirms the need to provide new online students a course-like orientation that operates as an interactive space, rather than simply offering a self-paced environment of video tutorials.

Keywords: online student orientation, online learning, degree completers

Increasing Connections to Increase Online Student Retention

Structured initiation of students to new educational settings is important for their academic success (Tinto, 1999 & 2006; Wozniak, Pizzica & Mahony, 2012). With online learners being the fastest growing segment of post-secondary students (U.S. Dept. of Education, 2014) and because one of the largest populations of potential college students is *degree-completing adults* (EAB, 2013) higher education practitioners must customize the introductory experience for this subpopulation of non-traditional learners (Home, 1998; Fairchild, 2003). Degree completers are students who accrued credits through a previous bachelor's or associate's program and seek to complete degree requirements; moreover,, when degree completion is sought at a different institution, they are also considered transfer students. Online degree completion programs are scaling in tandem (EAB, 2013) to this population, which suggests that adult learners are increasingly turning to online education in order to fit degree attainment into busy schedules, involving employment and domestic responsibilities, among others (Fairchild, 2003; Donaldson & Townsend, 2007; Ross-Gordon, 2011). Belonging to the nation's largest system of urban public education, working adult degree-completers make up the online bachelor's degree seeking population[1] at the City University of New York (CUNY) School of Professional Studies (SPS). CUNY SPS students are on average in their mid-thirties and transfer in almost the equivalent of an associate's degree, approximately 60 credits, if not a complete associate's degree (in some cases, even having already earned a bachelor's). Having attended one or multiple previous institutions, most students have some history in the CUNY system, with a minority having previously taken fully online courses. Possibly most significant about this population, the majority of students report being employed full time. These students must

[1] Bachelor's degree students must transfer a minimum of 24 credits and earn a minimum of 30 local credits.

acclimate to a new institution and a new learning environment/medium (online, within a specific learning management system), all within the context of major demands external those of the academy.

Social and Professional Connections

Evidence supports that lack of connections to peers, to faculty, and, more broadly, to the institution directly contributes to students' decisions to withdraw from university (Braxton & McClendon, 2001; MacKie, 2001; Tinto, 1999 & 2006). Participatory intellectual and personal communities provide access to information, resources, and support (Granovetter, 1973; Dawson, 2008) and these social networks foster stability and positive affect for students (Tinto, 1999, 2006). Because membership in small, like-minded groups has a strong influence on member behaviors (McPherson, 1998; Tsvetovat, 2011), we reasoned that comparable community behaviors exist in online education. Online interactions provide opportunities to develop student-institution relationships, as digital networks afford strong transactional and information-sharing behaviors (Milne, 2007, Brill & Park, 2008).

Adult learners are distinguished by their professional experience (Ross-Gordon, 2011) and online adult learners tend to be goal oriented and motivated by professional enrichment, seeking advancement in a given area or career changes (Fairchild, 2003; Howell, Williams & Lindsay, 2003). Connecting students to their intended profession or discipline can have a positive impact on students, academically and on personal levels (Folsom and Reardon, 2003). Career development courses can be associated with improved retention rates, and participation in these courses can also improve student's self-awareness and cognitive skills (Folsom and Reardon, 2003). Connection to a discipline can manifest in a number of ways, including networking students directly to practitioners who serve as mentors, and assisting students in developing a clear understanding of what professionals in that field do.

Purpose of study

With funding from the CUNY system office, a team comprised of academic leaders and institutional research sought to design an orientation experience that attended to the needs of online working adult degree-completer students. In addition to introducing students to the college (the whos and the ins and outs), the campus (online environment), the project incorporated best practices of traditional orientation – helping students to build social connections as a foundation for academic support – but enhanced it for the working adult population by bridging the social networks through discourse of professions and career. This enhanced approach to orientation foregrounds development of interpersonal and disciplinary connections designed to provide students with access to social supports for their established or developing professional identities. By emphasizing the social dimensions of orientation through structured activities designed to clarify students' vision of a post-degree professional self, situated within a chosen discipline or career, we aimed to foster a sense of connection and shared purpose to accelerate the development of strong social networks that promote student success.

The purpose of this study[2] was to improve student retention and performance outcomes by enhancing the social components of orientation for new online students to our bachelor degree completion programs. We looked to achieve this by structuring orientation as a short-term course and implementing elements of exemplary online course design, developing custom media content and devising "course" assignments designed to provide students the experience of

[2] This study was supported by the City University of New York Office of Academic Affairs Student Success Grant, awarded to "colleges to conduct rigorous evaluations of promising innovations designed to improve students' prospects for baccalaureate or associate degree attainment" (CUNY Office of Academic Affairs Request For Proposals, Nov 2012).

learning online, with comprehensive support from administration and peers, allowing new students to troubleshoot the online environment before launching into their curriculum. Additionally, students were given an opportunity to meet and interact with their peers, a peer mentor and faculty in their discipline (major) in order to begin building the connections necessary for support and success in learning and reinforcing their developing and established professional identities.

We hypothesized that an interactive orientation experience, one through which students actively engage with each other and have the intensive support and guidance of peer mentors and faculty, would foster the early establishment of social connections that would support the student in her success through the academic experience at the college. We also theorized that these connections would be most valuable to make at the disciplinary (program major) level, where

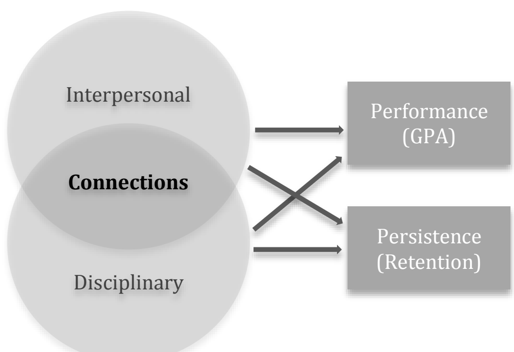


Figure 1 – Hypotheses of connections and student outcomes visualized

students could discuss their interests in the field and how the program curriculum would help in shaping their professional goals. The preliminary success of this model would be tested by looking at the outcomes after the students' first term, through academic performance, defined as grade point average (GPA), and persistence, defined as retention into a second term.

The study was designed as a controlled experiment, whereby the new student population was divided equally[3], half to receive the treatment of an enhanced orientation, while the second half, as the control group, completed the orientation without the intensive disciplinary interactions.

Study Design

Orientation was structured as a short-term course that would allow students to learn how to navigate the learning management system (LMS) of Blackboard Learn, participate in Discussion Board fora, complete graded assignments, and receive feedback. A syllabus detailed the activities allocated to each week (see Syllabus in Appendix A). The orientation site was further designed to collect data on students' participation in activities. Tracking was activated on all content, assignments were graded in the system with automated feedback for tests, and rubrics were used for manually graded assignments, including a reading and writing assignment[4] graded by General Education English faculty. The orientation site developers created new media to welcome the students, to aid students in navigating the site, and to educate them about CUNY SPS resources. Blackboard Inc. video tutorials were sourced for LMS-specific guidance. Orientation launched with a live webinar that was recorded and made available within the site. An MS Excel Time Management Instrument (see Appendix B) was developed and used as an activity, planning tool and introduction to spreadsheet software commonly used in quantitative

[3] Assignment was semi-random, whereby programs with small new populations had all students registered in the program enrolled in the enhanced orientation and larger programs had the population divided. This process was implemented to assure that there would be enough students present in the Discussion Board forum to meaningfully participate. In addition, students who registered for classes after the orientation launched were given access to the control site for completion of the requirements. Control in the analyses are defined as all students in the control site who were enrolled within the first week of orientation.

[4] Modeling after the University of Texas (Tough, 2014), motivational texts were selected, that emphasized lifelong learning and brain plasticity.

courses. Finally, an anonymous feedback survey concluded the orientation experience, after which, students could retrieve their earned Certificate of Completion.

Enhanced (treatment) orientation was designed with an additional content area, Groups by Major, where students could connect with their disciplinary classmates, peer mentor and faculty facilitator through a series of structured prompts discussing careers, goals, and curriculum (see sample prompts in Appendix C). Peer and faculty/professional mentors were hired to facilitate the discussions within disciplines and to answer questions about the new students' chosen major. Rubrics were created to assess the content of these discussions for analysis purposes, but posts were not graded.

To encourage and assess participation and interaction, an orientation facilitator was hired to welcome students, post regular/scheduled announcements and updates, and facilitate the large discussion boards, including a general Q&A area. In addition, biweekly data extracts were run to assess incomplete assignments and absence of participation, and customized messages were submitted to students in need of completing activities as well as reinforcing feedback for those who were on track with the syllabus requirements. In this way comprehensive outreach was established, encouraging students to complete the activities[5].

Methods

Our core hypothesis was that participation in the enhanced three-week new student orientation would facilitate strong initial connections with peers, faculty, and the chosen profession, which would in turn improve students' persistence and academic performance. Specifically, the following two sets of hypotheses were formulated:

[5] At the time of the study orientation was not an official prerequisite, and historically it has been difficult to motivate all students to complete all orientation components.

Hypotheses #1 and #2: Connecting students to one another and to faculty will increase first-term/first-year persistence and academic performance.

Hypotheses #3 and #4: Students' connection to their intended profession will increase first term/first-year persistence and academic performance.

Operationalized variables

To test formulated hypotheses we operationalized constructs in the following way: the two primary outcome variables of main interest were performance, a continuous variable measured as term GPA; and retention, a dichotomous variable denoting whether or not a student registered for courses the next term.

We performed content analysis of students' online posts on the group discussion board during the orientation period to measure connections to peers, faculty members and chosen profession. Each week students were asked to perform a number of activities that included reading program-specific content and writing and posting 250-300 word responses to prompts on the program discussion board. Specifically, during the first week students were asked to write a career narrative explaining what attracted them to their chosen field/profession. During the second week students were asked to write a response that connected careers to curriculum. Finally, during the third week of enhanced orientation, the students' task was to post to the program discussion board reflecting on the orientation experience and on participating in discussions about careers. Students were also asked to comment and to reply to at least two posts of their classmates each week to foster peer-to-peer engagement.

We considered several measures of the degree of students' involvement in enhanced orientation activities in our analyses. One variable was the number of narratives posted by each student on the discussion board, where higher value on the number of posts represents higher involvement in the activities. Additionally, the quality of posts was assessed using a rubric with four dimensions and a five-point likert scale: (1) completeness of the posts - whether all questions of the assignment were addressed in a post, (2) relevance to assigned question - this factor was different each week since prompts tackled different questions, (3) depth of posts - the quality of details and supporting arguments presented, (4) connection to personal experience - it was important for students to make connections between activities that they were performing and actual academic and career goals. These rubrics can be found in Appendix D.

In order to test our hypotheses regarding the influence of participation on term GPA and retention rate, we developed additional variables to assess students' degree of involvement in orientation activities: achievement of certificate[6] of completion (a binary variable designating whether or not a student completed all required tasks, and thus earned a certificate of completion), 2-week participation[7] (a binary variable indicating whether or not a student posted responses to prompts during the first two weeks), full orientation participation (a binary variable, whether students earned a certificate of completion, completed the time management tool, and posted responses to prompts during the first two weeks of orientation).

Content analysis results were also used to quantify students' connection to their intended profession. One version of the variable was defined as the sum of scores that students received for their narratives on week 1 and week 2. The following criteria were used to calculate the scores (responses were evaluated using 5-point Likert scale): Has clear and well-defined career

[6] Requirements for earning the Certificate of Completion included: a Blackboard Basics Quiz, posting to a *Challenges and Obstacles* Discussion Board forum, submission of Reading and Writing Assignment, reading and marking as read the Sexual Harassment Policy, and completion of Feedback Survey.
[7] We excluded the third week orientation results because topic of the prompt was not directly relevant to primary hypotheses. Additionally, we observed a strong response bias at the third week – those students that made a post that week expressed very positive attitude towards orientation, but there is no information as to why other students did not post narratives, if is it due to negative attitude, exhaustion of topic, or other unknown reasons.

goals; Understands the alignment of academic program with careers in the field; Relates the prompt to own goals, interests and expectations. Another form of this variable was defined as a sum of scores for the narratives on week 1 and 2 using the following criteria: Has clear and well-defined career goals; understands the alignment of academic program with careers in the field.

We operationalized connections with peers and faculty as the number and quality of students' responses to classmates, faculty members and peer mentors. We also calculated the number of feedback posts (without coding the content) that each student received from peers, faculty members and peer mentors, and used these variables as measures of interpersonal connections among students and faculty.

Sample

In our study we considered a cohort of entering first-time transfer students from Fall 2014. Out of 345 participants we used data from 216 students (Control group n=97, Experimental group n=119). Those who dropped registration (n=55) or registered so late that could not participate in orientation activities (n=82, there is an overlap with other excluded categories) were not considered in further analyses (Table 1).

Table 1. Frequencies breakdown by condition and retention

	Did not retain	Retained	Dropped registration	Total
Control	28	69	24	121
Experimental	36	83	23	142
Not NSO	8	15	0	23
Control (late registration)	15	36	8	59
Total	87	203	55	345

In evaluating the effect of enhanced orientation on term GPA, we also excluded students who withdrew from all courses, a result of a null GPA, because these participants did not have term GPA records. For GPA analyses the resulting sample size is n=201.

Demographics and educational background characteristics. The final sample of 216 participants was primarily comprised of females (n=159, 73.16%). Among those who reported their ethnicity White (n=63, 32.6%), Black (n=53, 27.5%), and Hispanic (n=48, 24.9%), ethnic groups were roughly equally presented in the sample. The resulting sample consisted of older students, whose age ranged from 19 to 63 years old with the mean age of 33.61 years old. In terms of educational background, students in the final sample were on average 3.7 years out of school with the average incoming GPA of 3.05. Over seventy percent (70.4%, n=152) of the sample studied at CUNY colleges in the past, and roughly half of students entered the program with an Associate degree (n=101, 46.8%). Combining data sourced from the Time Management tool and the Admissions intake form, 70% (n=109) of participants whose data were available were full-time employees with the median 45-hour work week. Every third student had no prior online education experience (n=48, 31.2%).

Results

Academic Performance Models

In our study we hypothesized that building strong initial connections with peer students and faculty will improve students' academic outcome. We were also interested in investigating which background characteristics may have potential impact on students' academic performance irrespective to treatment assignment. Academic performance in such models was defined as students' term GPA; therefore students who withdrew from all courses were excluded from further analyses due to having null term GPA. Academic performance models are discussed below.

Treatment effect. In order to evaluate treatment effect of enhanced orientation – experimental versus control group membership – on students' academic performance, and to make groups more comparable, we selected those students who demonstrated active participation in orientations by *Completed requirements and obtained certificate* dichotomous variable. The resulting samples consisted of 54 and 70 students in control and experimental groups respectively.

We performed analysis (ANCOVA) on active participants sample to evaluate treatment effect on academic performance after controlling for incoming GPA. Our results showed that treatment effect was not significant ($F(1,121)=.787$, $p=.377$), suggesting that the two groups of active participants did not demonstrate statistically significant difference in term GPA adjusted for educational background.

Although we did not obtain direct proof of strong positive effect of enhanced orientation on students' academic performance, some of its aspects appeared to have a significant effect and explain a portion of term GPA scores' variance. Additionally, we found that some background characteristics have predictive power. These models are discussed below.

Interactions with students/faculty/peer mentors. As mentioned earlier, we operationalized academic performance as term GPA and used it as dependent variable in a set of multiple regression models with enhanced orientation variables as predictors.

Among orientation variables that were obtained as a result of content analysis, *Feedback from students* – number of responses that students receive from peers – appears to be a significant predictor of term GPA ($b = .158$, $p=.018$, $F(2, 67)=3.259$, adjusted $R^2=.06$) after controlling for incoming GPA. These findings suggest existence of positive effect of interactions among students, defined as number of feedback messages that students receive from classmates, on academic outcome.

The results did not change when we looked at the effect of *Feedback from students* on term GPA after partialling out the effect of potentially influential variables – *financial aid* and *online experience*. Despite reducing the power of the test due to decreased sample size, the model appeared to be significant (b = .145, p=.028, $F(3, 51)=2.93$, adjusted $R^2=.10$).

Interestingly, although *Feedback from faculty member* or *Feedback from peer mentors* do not have an effect individually, when we fitted a more general version of the model with the *Total number of feedback messages* as a predictor of academic performance alongside with influential covariates, this composite score also demonstrated predictive power (b = .085, p=.028, $F(3, 51)=2.95$, p=.042, adjusted $R^2=.10$). Based on the results we can conclude that greater support that students receive, defined as number of feedbacks from other students, faculty, and peer mentors combined, has a positive effect on academic performance.

Overall active participation in orientation. We hypothesized that active participation in orientation, measured as number of responses to prompts written and posted by students (0 – minimum, 3 – maximum) will have positive effect on their academic performance expressed as term GPA. Our findings suggest that *Total threads count* variable is indeed a significant predictor of term GPA (b = .23, p=.046, $F(2, 109)=3.58$, $R^2=.04$). According to obtained results, the more actively students participate in orientation, the higher term GPA such students tend to have, even after controlling for incoming GPA.

In our analyses we also operationalized students' involvement in orientation as a number of fulfilled requirements (6 is the maximum) for the whole population. Our findings suggest that the number of completed requirements is a significant predictor of academic performance at the end of the term, even after controlling for incoming GPA (b = .135, p=.008, $F(2, 198)$=7.19, adjusted R^2=.06). In other words, students that demonstrate higher involvement in orientation activities defined as the number of completed requirements, tend to have higher GPA at the end of semester, even after partialling out effect of incoming GPA.

Effect of reading motivational texts. Because our target population consisted predominantly of older students who have been out of school for several years, one of the goals of orientation was to address students' potential anxieties and concerns about their academic performance and ability to learn (Tough, 2014). As a part of orientation all enrolled students regardless their treatment assignment were asked to read motivational texts about brain plasticity and lifelong learning, and to write a short reaction paper whose goal was to ensure implementation of reading assignment.

We performed analysis of covariance (ANCOVA) to evaluate effect of reading motivational texts on term GPA. Our findings suggest existence of positive effect of the reading assignment on academic outcome. According to our results, students who completed reading assignment tend to have better academic performance[8] compared to those who did not, even after controlling for incoming GPA ($F(1, 197)$=8.78, p=.003, adjusted R^2=.067).

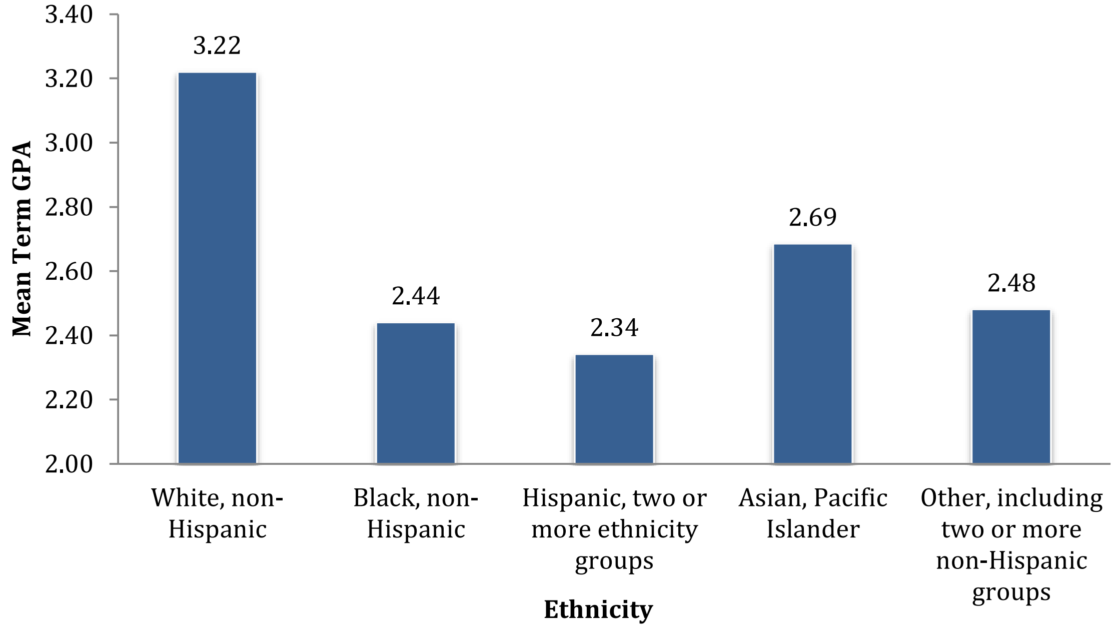
[8] We fitted logistic regression to evaluate effect of assignment completion on students' persistence, but our findings did not support this hypothesis.

Time management. The Time Management (TM) instrument was included as part of a discussion board assignment for both groups. Given that 70% of the sample self-identified as full-time employees at the time of registration[9], and a median workload of 45 hours per week was identified through the TM instrument, we investigated if variables from the TM instrument have predictive power on students' performance. Our results demonstrated that the more hours students are planning to allocate for each class in their weekly schedule, the higher term GPA these students tend to have (b = .041, p=.037, F(1, 124)=4.440, R^2=.04). These findings underline the importance of efficient time management and realistic expectations in the academic success of online students.

Ethnicity gap. One of the important concerns often discussed in the educational research literature is students' ethnicity gap in academic performance (e.g. Jencks, C., & Phillips, M., 2011). To address the issue we ran an analysis of variance (ANOVA) to compare term GPA of different ethnic groups, and obtained significant results (F(4, 174)=2.977, p=.021, $\eta^2 = 0.10$). A Post-hoc Tukey HSD pairwise comparison of term GPA revealed statistically significant difference in the average academic performance between White and Black groups as well as between White and Hispanic students. In both situations White non-Hispanic students outperformed other ethnic groups (Figure 1).

[9] As a part of enrollment process incoming students are asked to complete an intake form, a survey with questions regarding background and time management plans, within which employment status is identified.

Figure 1. Term GPA by Ethnicity



Retention Models

In our study we hypothesized that connecting students to one another and to faculty will increase first-term/first-year persistence. Additionally, we looked at background characteristics that may have potential impact on students' retention regardless treatment group. Retention in such models was operationalized as a dichotomous variable, denoting whether or not a student registered for the next term. Persistence models are discussed below.

Treatment effect. To evaluate effect of treatment – experimental versus control group membership – on students' retention, and to make a valid comparison between the groups, we selected those students who demonstrated active participation in regular or enhanced orientation. For such purposes we considered participation to be active if students completed requirement and received certificate of completion (N=61 in control and N=75 in experimental group).

We fitted logistic regression using a sample of active participants to evaluate treatment effect on retention after controlling for demographic characteristics, such as age, gender, and race. Our results showed that the treatment effect was not significant (b=.391, $\chi^2(1)$= .806, p=.369, odds ratio of reenrollment=1.478), suggesting that group membership did not account for a significant portion of variation of persistence after controlling for demographic characteristics.

We employed strategies similar to academic performance models, and despite obtaining an insignificant treatment effect, we examined which aspects of enhanced orientation as well as students' background characteristics could have a predictive power for retention. These models are described below.

Interactions with students/faculty/peer mentors. According to the second formulated hypothesis, connecting students to one another and to faculty will improve retention rate. We defined *retention* as a fact of registration to the next semester, and used it as dependent variable in a set of logistic regression models with enhanced orientation variables as predictors.

Consistent with the term GPA models, among a set of orientation variables that were obtained as a result of content analysis, *Feedback from students,* measured as a number of feedbacks that students receive from peers, appeared to be a significant predictor of retention (b=.247, $\chi^2(1)$= 4.179, p=.041, odds ratio of reenrollment =1.281, Nagelkerke R^2=.27), after controlling for the effect of *age*, *gender* and *ethnicity* characteristics. These findings suggest that with the increased number of interactions among students defined as a number of feedback messages from peers, the probability of retention at the program increases.

Background characteristics for all student population. As a part of our study we investigated the impact of several background characteristics that potentially influence students' retention regardless treatment group. This sample, combining control and treatment groups, was comprised of 136 students.

Multiple logistic regression with demographic variables of *gender*, *race* and *age* did not demonstrate a good fit and was not significant. In a logistic regression model with academic background variables, including *incoming GPA*, *highest degree attained*, *number of colleges attended*, whether a student has ever *studied at CUNY colleges* before or has *previously taken online classes*, none of the mentioned predictors had an effect on retention probability.

Because our sample primarily consisted of full-time employees with domestic and other commitments, we followed the same logic as with term GPA models and investigated predictive power of *employment* and *hours devoted to job per week* variables in retention models. Despite *employment status* not proving to be a good predictor of persistence, *hours devoted to job per week* significantly predicted retention (b= -.032, $\chi^2(1)$= 3.905, p=.048, odds ratio of reenrollment =.969, Nagelkerke R^2=.059), meaning that greater hours devoted to work reduces the probability of students' retention at the program. Another significant factor that plays a role in students' persistence is being a *financial aid*[10] recipient (b = .836, $\chi^2(1)$= 5.173, p=.023, odds ratio of retention =2.369, Nagelkerke R^2=.054). According to these results, students that receive any type of financial aid have higher probability to register for the next term compared to those

[10] Financial aid recipient was operationalized for this analysis as receiving: TAP and Veteran state aid, Pell, SEOG and loan federal aid, or waiver.

students who do not. These findings underline the importance of efficient time management and the role of financial components in students' persistence.

Connection to Profession Models

According to hypotheses #3 and #4, students' connection to their intended profession will have a positive effect on first term/first-year persistence and academic performance. To test these hypotheses we defined connection to the intended profession for the experimental group through variables that reflect quality of the narratives posted by students during the first and the second weeks of orientation (see methods section for description and appendix for rubrics): *sum of scores W1 (criteria #2 and #3) and W2 (#2 and #3)* and *sum of scores W1 (criteria #2) and W2 (#2).*

We fitted a set of logistic regression models with retention outcome as a dependent variable, and used described orientation variables as predictors, and obtained significant results (b= .130, $\chi^2(1)$= 3.958, p=.047, odds ratio of retention =1.136, Nagelkerke R^2=.089, and b=.250, $\chi^2(1)$= 3.932, p=.047, odds ratio of retention =1.284, Nagelkerke R^2=.089 respectively). There results suggest that students with better connection to a future profession, defined as a quality of responses to career reflection prompts, demonstrate higher persistence.

We fitted a linear regression model to test the hypothesis that students' connection to their intended profession increases academic performance measured as term GPA. According to the results, the model does not fit well to the data, and does not account for significant portion of term GPA variance.

Discussion

> *What surprised me about the orientation and assignments is that it was not just designed to acclimate me to the tools, but really helped me to begin to flesh out*

my philosophical approach to my career choices. All of the assignments were very relevant and helpful both practically and professionally. It really was not what I was expecting at all - it was much better! Thank you and I look forward to a productive and successful semester.

(Anonymous student feedback, August 2014)

The results of this study indicate a positive relationship between active and repeated engagement among new students during the orientation process and first term academic performance as well as persistence. These results support our hypotheses that interpersonal connections are beneficial to the online learning process. Discussion board participation is a ubiquitous requirement of online learning at the College. By acclimating new students to this practice before courses even commence, we are establishing a foundation for the necessary habits of learning online.

The hypotheses that disciplinary connections would also have a positive effect on degree persistence was also supported by the evidence in this study, although not confirmed for academic performance. The measures for these tests were about the quality of the posts made by students regarding professions. From the aforementioned results of a positive relationship in the quantity of posts to and from the student in combination with the positive relationship between professional clarity and retention, we see that the *socio-disciplinary interactions* are valuable in both quantity and quality. This suggests that students who are engaged and communicate with clarity their interests and thoughts perform better and are better retained. Fostering these habits of a “good” online student is important, and orientation affords students the opportunity to work through the dynamic of the online participation sphere.

Having observed positive correlations between engagement and student outcomes in academic performance and retention, the study's results are being used to inform the training process of peer mentors for orientation facilitation. Before the orientation period launches each term, peer mentors meet with the orientation team, consisting of an academic director, the orientation facilitator and institutional research, to review responsibilities and expectations. The evidence of this study has taught us to guide mentors towards a high touch approach: not only doing outreach to individual students, but facilitating and encouraging students to reach out to each other.

The finding that time management is crucial for successful online learning, both in working through full-time employment and allotting an appropriate amount of time for studies, can be addressed by the team of peer mentors, academic advisement and in the orientation site design. Peer mentors, as successful advanced students or recent graduates, are being trained to advise incoming students about how to manage the multiple responsibilities of adult degree completers. In addition to the tips offered by peer mentors, the time management activity will be refined to more specifically address balancing employment and studies. Academic advisement is also being integrated into the time management forum, where advisors can work with their appointees, one-on-one, using the Time Management instrument to discuss how much time is required for studies within the summary the students present in the tool.

Some additional modifications have been made to the orientation model studied here as a result of feedback provided by the students, the peer mentors, and the orientation facilitator: orientation has been restructured to a 2-week period, additional content was developed to help students better understand online library resources and academic integrity[11], a second, closing,

[11] A component for Title IX compliance was also added.

webinar was added to complement the welcome webinar, moving content from the original webinar about "What to do on the first day of classes" to the week classes begin.

Additional next steps at the college will include tracking out student performance and persistence to the 1-year mark for the study cohort. In this work we will look to see if effects hold long-term for student success and reenrollment patterns. We will also code the data for a complete new student cohort, whereby we can see if the effects show to be stronger with a larger sample. Another area for future study is the overall structure of our online courses. We will investigate ways to make the workload manageable for employed adult learners while retaining rigour and achieving course and program learning outcomes.

The study demonstrated that the more activities completed during orientation the better academic outcomes were seen at the end of the first term. Recommendations will be made by the orientation team, working in conjunction with the admissions leadership, to mandate completion of orientation for all new students as well as standardization and completion of the admissions intake survey. By engaging all students early, the performance and persistence outcomes of the overall population should be affected, as the data indicated that the quantity of interaction has positive effects. One of the limitations to this study was the sample size, whereby group selection for analyses reduced the number of student records to evaluate. This may be the cause for the size effect of some of the results not being as strong as expected. We believe that with comprehensive, mandatory, implementation more students will reap the benefits of participating in orientation, and the consequent larger population will produce more data to more deeply evaluate the strength of the effects already evinced.

As the University seeks to grow online education, the model for orientation developed at the College can be presented to other campuses. The presentation would include a full description of the resources, timelines for implementation, the syllabus and a template of the site. The findings of this study can also be shared in the broader online higher education community. The analyses support the importance of a dynamic facilitation of online orientation for adult degree completers; an orientation that encourages interaction between students and addresses the development of professional identities and effective time management. While many online programs place students into a self-paced orientation of tutorials and activities, this orientation model structures activities synonymous with a course, so that students are moving forward and learning together. Moreover, interpersonal and disciplinary connections support the new student to online learning, acclimating her to her new social and academic milieu, as well as to the practice that makes for a successful online learner.

Acknowledgments

The authors would like to thank Paul Russo for his work in the development of the model and study as part of the orientation team; Angela Francis for her facilitation of orientation; the faculty and peer mentors; the CUNY SPS Departments of Information Technology, Student Services and Admissions. A special acknowledgement is also made to the CUNY Office of Academic Affairs for supporting this study.

References

Anderson, T., Poelhuber, B. & McKerlich, R. (2010). Self-paced learners meet social software: An exploration of learners′ attitudes, expectations and experience. *Online Journal od Distance Learning Administration*, 13(3). Retrieved from http://www.westga.edu/~distance/ojdla/Fall133/anderson_poellhuber_mcKerlich133.html

Braxton, J. & McClendon, S. (2001). The fostering of social integration and retention through institutional practice. *Journal of College Student Retention*, 3(1): 57- 70.

Brill, J.M. & Park, Y. (2008). Facilitating engaged learning in the interaction age. *International Journal of Teaching and Learning in Higher Education*, 20(1): 70-8.

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology and Society*, 11(3): 224–38.

Donaldson, J. F. & Townsend, B. K. (2007). Higher education journals' discourse about adult undergraduate students. *The Journal of Higher Education*, 78(1): 27-50.

Education Advisory Board. (2013). *Future Students, Future Revenues: Creating a Principled and Sustainable Enrollment Strategy*. Washington, DC: Author.

Fairchild, E. E. (2003). Multiple roles of adult learners. *New Directions for Student Services*, 102: 11-16.

Folsom, B. & Reardon, R. (2003). College career courses: Design and accountability. *Journal of Career Assessment*, 11(4): 421-50.

Granovetter, M. S. (1976). The strength of weak ties. *American Journal of Sociology*, 78(6): 1360-80.

Henseler, J., Ringle C., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20: 277-319.

Hensley, L. G. & Kinser, K. (2001). Rethinking adult learner persistence: Implications for counselors. *Adultspan Journal*, 3(2): 88-100.

Holder, B. (2007). An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education online programs. *The Internet and Higher Education*, 10(4): 245-60.

Home, A. M. (1998). Predicting role conflict, overload and contagion in adult women university students with families and jobs. *Adult Education Quarterly*, 48(2), 85-98.

Howell, S. L., Williams, P. B., & Lindsay, N. K. (2003). Thirty-two trends affecting distance education: An informed foundation for strategic planning. *Online Journal of Distance Learning Administration*, 6(3). Retrieved from http://www.westga.edu/~distance/ojdla/fall63/howell63.html

Jencks, C., & Phillips, M. (2011). *The Black-White test score gap.* Washington, D.C.: Brookings Institution Press.

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers and Education,* 48: 185-204.

MacKie, S.B. (2001). Jumping the hurdles - Undergraduate student withdrawal behavior. *Innovations in Education and Teaching International*, 38(3): 265-76.

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations. *American Sociological Review*, 52(3): 370-79.

Milne, A. J. (2007). Entering the interaction age today. *Educause Review*, 42(1): 12-31.

Morris, L. V., Wu, S., & Finnegan, C. L. (2005). Predicting Retention in Online General Education. *American Journal of Distance Education*, 19(1): 23-36.

Tinto, V. (1999). Taking retention seriously: Rethinking the first year. *NACADA Journal*, 19(2): 5-9.

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

Tough, P. (2014, May 18). Who gets to graduate? *New York Times Magazine*, pp. MM26. Retrieved from http://www.nytimes.com/2014/05/18/magazine/who-gets-to-graduate.html

Tsvetovat, M. & Kouznetsov, A. (2011). *Social network analysis for start-ups: Finding connections on the social web*. Sebastopol, CA: O'Reilly Media.

U.S. Department of Education, National Center for Education Statistics. (2014) 2003-04, 2007-08, and 2011-12 National Postsecondary Student Aid Study [Data File] (NPSAS:04, NPSAS:08, and NPSAS:12).

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. Washington, D.C..

Wozniak, H., Pizzica, J., & Mahony, M.J. (2012). Design-based research principles for student orientation to online study: Capturing the lessons learnt. *Australasian Journal of Educational Technology*, 28(5): 896-911.

Appendix A
New Student Orientation (NSO) Syllabus, Fall 2014

CUNY The City University of New York | School of Professional Studies

ORIENTATION SYLLABUS

COURSE DESCRIPTION
This three-week workshop prepares transfer students to successfully navigate the online learning environment. Using the University's Blackboard course management system, participants will learn about the mechanics of coming back to school: time management, developing good study habits, developing the "reading and writing muscle" needed for online work, and other techniques of successful online learners. Focusing on developing interpersonal connections in an online environment, students meet and interact with faculty and future classmates in their programs; practice and receive feedback on college-level reading and writing assignments; begin career planning, and learn about useful services and resources at the CUNY School of Professional Studies.

ORIENTATION LEARNING OUTCOMES
Students will:

- Understand and be able to use all essential features of the Blackboard online course management system
- Create a time management plan based on current course schedule and life demands
- Read academic texts and demonstrate comprehension through evaluated written assignments
- Demonstrate a practical understanding of academic integrity
- Learn about useful services and resources at the CUNY School of Professional Studies, including academic and career advisement, prior learning assessment, ePortfolio, tutoring, and social media
- Connect career goals and learning expectations to academic discipline
- Establish social connections (community and small group) to other students through discussions
- Build social connections to faculty and industry advisors by discussing current topics and issues of interest in the discipline

REQUIRED TEXTS AND READINGS

Required texts are available in electronic formats either accessible on, or linked to, the class website in Blackboard 9.1, the course management system licensed by CUNY. For example, the Purdue Online Writing Laboratory (OWL) will be referenced for information on APA Style Citation and the CUNY Information Literacy Tutorials for an introduction to academic research online.

WEEK	OBJECTIVES	READING/LECTURE	ACTIVITIES/ASSIGNMENTS
1	• Introduction to SPS and to Orientation; • Learn about and practice using Blackboard Course Management System; • Learn time management strategies; • Meet peers, faculty and professionals in your field of study; • Connect career goals and expectations to academic discipline.	• Watch Dean's welcome video • Watch Getting Started video • Review Blackboard Tutorials • Watch time management video & read tips • Read about career opportunities for your major	• Participate in Welcome Webinar • Complete Blackboard Quiz; • Send an email to your group's peer or professional mentor via the Orientation Blackboard site; • Post and Comment on Introductions Discussion Board; • Post and Comment on Overcoming Challenges Discussion Board; • Post to Group Discussion Board Forum: Career Narrative Part I;
2	• Practice and receive feedback in academic reading and writing; • Learn to use online collaboration tools • Meet other students in your program • Understand academic integrity and avoiding plagiarism;	• Read information about academic integrity • Read articles "Grow Your Brain" and "Personal Best" • View Writing Evaluation Rubric • Read information about academic integrity • Watch Collaborating Online video • Review Major Curriculum Listing from SPS Bulletin	• Post Change essay in response to the readings on the DB and submit as an assignment; • Post and Comment on Group Discussion Board Forum: Connecting Careers to Curriculum
3	• Reflect on orientation experience and next steps. Are you ready? • Learn about SPS resources available to you	• Review CUNY SPS Student Policies • Watch FAQ video • Read about Credit for Prior Learning • Watch ePortfolio video	• Download and complete Time Management Tool • Post and Comment on Time Management Charts and Plan Reflection Discussion Forum • Post and Comment on Discussion Board: Jobs and Career Tracks in your Field • Complete New Student Orientation Feedback
		• Watch Social Media Tour video • Review Info Literacy Tutorials & video	Survey • Career narrative reflection: have your thoughts about your initial career narrative changed? • Download your New Student Orientation Completion Certificate

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see:http://sps.cuny.edu/student_services/disabilityservices.html

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

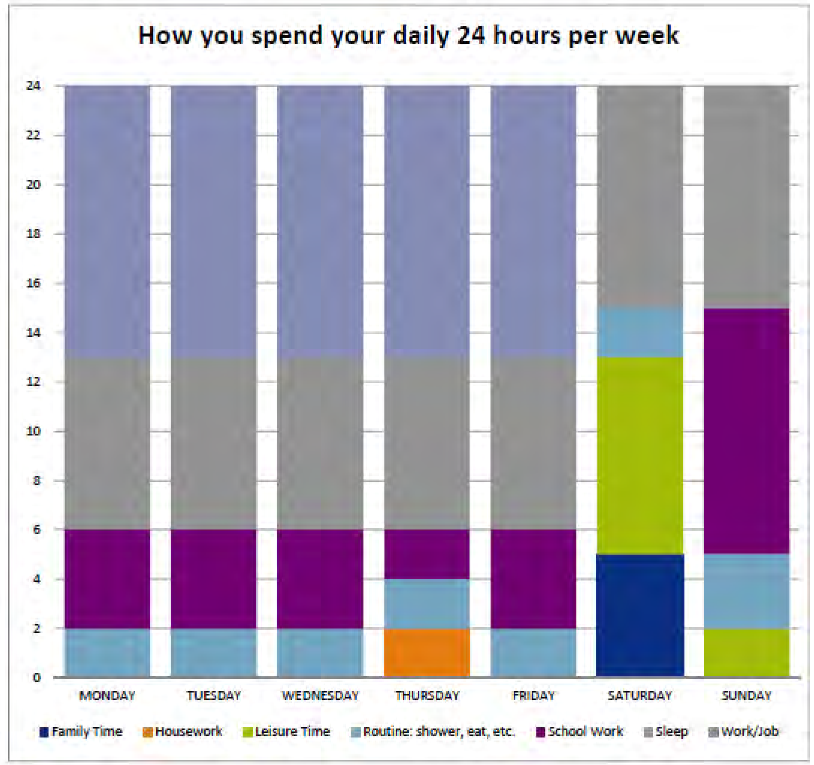
STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services: http://sps.cuny.edu/student_resources/

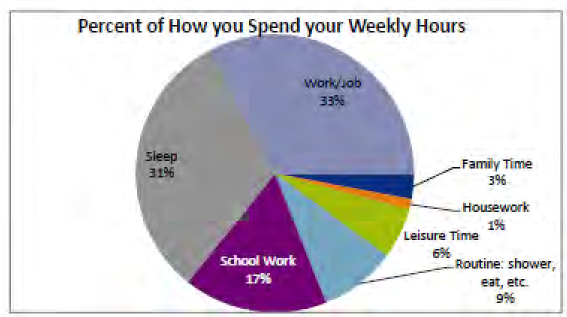
Appendix B
Sample Time Management Instrument

Time Management Instrument

HOURS	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
12:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
1:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
2:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
3:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
4:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
5:00 AM	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep	Sleep
6:00 AM	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Sleep	Sleep
7:00 AM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Sleep	Sleep
8:00 AM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Sleep	Routine: shower, eat, etc.
9:00 AM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Routine: shower, eat, etc.	School Work
10:00 AM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Leisure Time	School Work
11:00 AM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Leisure Time	School Work
12:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Leisure Time	School Work
1:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Leisure Time	School Work
2:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Leisure Time	Routine: shower, eat, etc.
3:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Family Time	School Work
4:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Family Time	School Work
5:00 PM	Work/Job	Work/Job	Work/Job	Work/Job	Work/Job	Family Time	School Work
6:00 PM	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Routine: shower, eat, etc.	Family Time	School Work
7:00 PM	School Work	School Work	School Work	Housework	School Work	Family Time	School Work
8:00 PM	School Work	School Work	School Work	Housework	School Work	Routine: shower, eat, etc.	Routine: shower, eat, etc.
9:00 PM	School Work	School Work	School Work	School Work	School Work	Leisure Time	Leisure Time
10:00 PM	School Work	School Work	School Work	School Work	School Work	Leisure Time	Leisure Time
11:00 PM	Sleep	Sleep	Sleep	Sleep	Sleep	Leisure Time	Sleep



Category	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY	Total Hours in
Family Time	0	0	0	0	0	5	0	5
Housework	0	0	0	2	0	0	0	2
Leisure Time	0	0	0	0	0	8	2	10
Routine: shower,	2	2	2	2	2	2	3	15
School Work	4	4	4	2	4	0	10	28
Sleep	7	7	7	7	7	9	9	53
Work/Job	11	11	11	11	11	0	0	55
Total Hours a	24	24	24	24	24	24	24	



Appendix C
Sample Groups by Major Prompts

Sociology	Use this forum to ask your mentors any questions you have about your program. To ask a question, click on the title of the forum to the left. Then click add new thread to make your post.
Week 1: Career Narrative: Part I	You're making a serious investment of time and resources to finish your degree, and in part betting your future on education. Good choice! Before the semester gets started, let's take some time to chart out your career. We'll look not just at current jobs in the field, but also emerging trends. A college degree is all about *possibility* and so is this discussion. **Step One**: To begin this assignment, read Pathways to Job Satisfaction.pdf **Step Two**: Sociology majors get jobs in a number of different fields. View this list of Job Titles for Sociology Majors.pdf (courtesy of Univ. Notre Dame) and start thinking about what jobs are available for you. Which positions interest you the most? Before you answer, spend an hour on Indeed.com searching for and reading job postings for the titles that you want. Try to get a sense of what these jobs are like and what skills you'll need to do well. For some sociology-based jobs, you may benefit from earning a master's level degree in sociology. What is it specifically that interests you about these positions? Is it a passion for the field, working with people, job security? Something else? Tell us which titles you're considering and why. Did your choices change at all as a result of your job research? How so? You have so many options for the future. It would be a missed opportunity not to map out the ones that interest you the most before you start taking classes! **Step Three**: Write a post of 250-300 words answering the above questions, as you map out your career narrative. To participate in this and all of the discussions, click on the title of the forum to the left. Then click add new thread to make your original post. **Step Four:** After you have submitted your post, read through your classmates' posts and respond to at least two of them. Note that you must first post your own submission before you can see others' posts.
Week 2: Connecting Careers to Curriculum	The Career Narrative assignment had you produce a thoughtful career plan. Now you want to read through the course catalog and identify the courses you plan to take from your major and explain how these courses will help you get the job you want (identified in Week 1). **Step One**: Review the courses in your major. **Step Two**: In a 250-300 word post to this DB forum, list at least five courses from your major's curriculum that will help you get (prepare for) that job. Explain the following in your post: How does the course description fit with the job description/the skillset required for the job? Do the course options fulfill your expectations for this degree? Are there courses that SPS doesn't offer that you think would be helpful? If so, what are they? As you write your response to the questions above, you want to make clear connections between the curriculum offered at CUNY SPS and the career path you have mapped out. Also feel free to ask your mentors questions about the courses. **Step Three:** After you submit your post, read through your classmates' posts and respond to at least two of them.

Appendix D
Enhanced NSO Careers Discussion Board Rubrics

Week #1 (Career Narrative Prompt):

Criteria / Score	**1 (Strongly Disagree)**	**2 (Disagree)**	**3 (Neither Agree Nor Disagree)**	**4 (Agree)**	**5 (Strongly Agree)**
Addresses the prompt /provides complete answer to the prompt	Does not address the assignment; the answer is irrelevant to the prompt.	Somewhat addresses the assignment, answers 25% of questions.	Addresses the part of the assignment, answers 50% of questions.	Addresses the assignment, answers 75% of questions.	Addresses the assignment completely, answers all questions.
Has clear and well-defined career goals	Doesn't have defined career goals and general understanding of the field.	Doesn't have defined career goals and has general understanding of the field.	Has somewhat defined career goals and general understanding of the field.	Has somewhat clearly defined career goals and good understanding of the field.	Has clearly defined career goals and deep understanding of the field.
Relates the prompt to own goals, interests and expectations	Does not relate the prompt to own interests and expectations.	Shows some consideration of how the prompt relates to own interests and situation.	Shows some thinking and reflection of how the prompt relates to own interests and situation.	Shows good thinking and reflection of how the prompt relates to own interests and situation.	Shows superior thinking and deep reflection of how the prompt relates to own interests and situation.
Demonstrates rationale /considers alternatives	Shows no rationale or supporting evidence. Speaks only in generalities.	Does not consider alternatives, provides minimal rationale for thinking or supporting evidence. Primarily speaks in generalities.	Considers alternatives, provides few supporting evidence or examples for rationale. Speaks in generalities, provides few details.	Considers alternatives, provides moderate supporting evidence or examples for rationale. The answer is primarily specific.	Demonstrates consideration of alternatives and supports thinking with solid evidence and/or examples. The answer is very specific.

Week #2 (Connecting Careers with Curriculum Prompt):

Criteria / Score	1 (Strongly Disagree)	2 (Disagree)	3 (Neither Agree Nor Disagree)	4 (Agree)	5 (Strongly Agree)
Addresses the prompt /provides complete answer to the prompt	Does not address the assignment; the answer is irrelevant to the prompt.	Somewhat addresses the assignment, answers 25% of questions.	Addresses the part of the assignment, answers 50% of questions.	Addresses the assignment, answers 75% of questions.	Addresses the assignment completely, answers all questions.
Understands the alignment of academic program with careers in the field	Doesn't demonstrate understanding of the connection of academic program with the career goals and careers in the field; Doesn't provide any explanation for course selection, doesn't make any connection with the career goals/interests.	Demonstrates weak understanding of the alignment of academic program with career goals and careers in the field; Provides unclear explanation for course selection/ doesn't make clear connection with the career goals/interests.	Demonstrates moderate understanding of the alignment of academic program with career goals and careers in the field; Doesn't provide well-thought explanation for course selection/ makes few connections with the career goals/interests.	Demonstrates somewhat clear understanding of the alignment of academic program with career goals and careers in the field; Explains course selection well, doesn't make clear connection with career goals/interests.	Demonstrates clear understanding of the alignment of academic program with career goals and careers in the field; Clearly explains course selection, makes connection with the career goals/interests.
Relates the prompt to own goals, interests and expectations	Does not relate the prompt to own interests and expectations.	Shows some consideration of how the prompt relates to own interests and situation.	Shows some thinking and reflection of how the prompt relates to own interests and situation.	Shows good thinking and reflection of how the prompt relates to own interests and situation.	Shows superior thinking and deep reflection of how the prompt relates to own interests and situation.

Demonstrates rationale /considers alternatives	Shows no rationale or supporting evidence. Speaks only in generalities.	Does not consider alternatives, provides minimal rationale for thinking or supporting evidence. Primarily speaks in generalities.	Considers alternatives, provides few supporting evidence or examples for rationale. Speaks in generalities, provides few details.	Considers alternatives, provides moderate supporting evidence or examples for rationale. The answer is primarily specific.	Demonstrates consideration of alternatives and supports thinking with solid evidence and/or examples. The answer is very specific.

Using Subsequent Course Performance to Evaluate the Effect of Differentiated Instruction in Math

One College's Formative Experience

Paper Presented to

The Northeast Association of Institutional Researchers (NEAIR)

November 2, 2015

Burlington, VT

By

Jerome J. Dean

Director of Institutional Research

Mount Ida College

Newton, MA

Introduction

Differentiated instruction has its origins in elementary and secondary education. Washburn (1953) documents efforts at the elementary level going back as far as 1889 in the United States of individual instructors recognizing distinctions between students' readiness to master concepts and attempting to address those distinctions within the instructional process using what amounted to differentiated instructional approaches. These early efforts at the elementary school level were well in advance of the advocacy of practitioners like Carol Ann Tomlinson, whose calls for greater use of differentiated instructional techniques (Tomlinson, 1995) came in the wake of the 1990 Americans with Disabilities Act and the movement toward "inclusion" of special needs children into the traditional K-12 classroom over the course of the 1990s. As a consequence, there is some evidence of the positive effect of differentiated instruction on learning gain at the K-12 level (Subban, 2006; Lightweis, 2013; Dosch and Zidon, 2014).[1]

In contrast, the evidence of positive effect is more limited at the post-secondary level. Dosch and Zidon (2014: 345) note, "[a]t the college level, even fewer studies exist regarding differentiation" for four reasons: "(a) class sizes are typically quite large; (b) the number of contact hours with students is minimal; (c) designing several ways to assess students is time consuming and challenging for professors who, in addition to teaching, have research and service obligations; and, finally, (d) ethical issues such as fairness in grading can be controversial."[2] The limited number of studies appears to be a consequence of a more limited use of

[1] Each of these articles cite a number of studies purporting to show positive effects of differentiated instruction on students, including positive learning gains.

[2] Dosch and Zidon (2014) are citing findings from Ernst, H. R., & Ernst, T. L. (2005). The promise and pitfalls of differentiated instruction for undergraduate political science courses: Student and instructor impressions of an unconventional teaching strategy. Journal of Political Science Education, 1(1), 39-59.

differentiated instruction techniques at the college level. Furthermore, of the five existing studies identified in Dosch and Zidon (2014) only 2 were clearly designed to identify learning gain as the main indicator.[3] The Dosch and Zidon (2014) study itself represents a third study where learning gain was the main evaluative measure. While the first 2 studies they cite used pre- and post-tests as the indicator of learning gain, the Dosch and Zidon study used student performance on a series of assessments in a psychology course. Two of the three studies were designed as experiments and one was a quasi-experiment.

This paper contributes to the literature regarding the learning gain associated with differentiated instruction in mathematics. Specifically, this study examines the impact of differentiated instruction in a remedial math course, at a small New England Liberal Arts College, on completion of the remedial course itself and subsequent remedial student performance in two college level courses. This study was not designed as an experiment but takes advantage of the natural experiment associated with the College's abrupt shift from a two-course remedial math regimen to a one-course regimen using a differentiated instructional approach.

What is Differentiated Instruction?

Generally differentiated instruction recognizes that "[t]odays classrooms are filled with diverse learners who differ not only culturally and linguistically but also in their cognitive abilities,

[3] The first study cited [Chamberlin, M., & Powers, R. (2010). The promise of differentiated instruction for enhancing the mathematical understandings of college students. Teaching Mathematics and Its Applications, 29, 113-139.] used a pre-and post-test in mathematic. The second study cited [Tulbure, C. (2011). Differentiate instruction for preservice teachers: An experimental investigation. Procedia—Social and Behavioral Sciences, 30, 448-452.] also used a pre- and post-test in science. Note that the only study of learning gain cited in Lightweis (2013) is the same Chamberlain and Powers (2010) study cited by Dosch and Zidon (2014).

background knowledge, and learning preferences (Huebner, 2010)." The approach requires a pre-assessment of a student's ability to perform/understand a set of course concepts or tasks and an assessment of student abilities on those tasks/concepts after a set period of instruction. Ideally instruction should be designed to "meet each student's individual learning needs," and in doing so, instruction for any individual student at any given point in time should be focused on those areas where the student has not demonstrated prior competence or even mid-instruction competence. Assessment of competence must be an ongoing component of the instructional process (Tomlinson, 1995; Hall, 2002; Levy, 2008; Huebner, 2010). Therefore, classrooms and instructional time should be tailored to maximize the amount of time that students spend on those competencies with which they are having the most difficulty.

The College's Approach to Differentiated Math Instruction

Before the introduction of the College's new differentiated developmental math course (MA093) in the spring semester of 2012, all developmental math courses at the College were taught in the traditional manner. This traditional developmental course typically involved a lecture that met 2 or three times a week for a total of 2.5 hours, a series of quizzes on different sections of course content that were part of the students' final grades, a set of 3 exams included in the final grade and the opportunity for students to meet one on one with the course instructor during office hours (or by appointment) for additional help. There were two levels of developmental math: MA090 which focused on basic mathematical concepts and computational skills, and MA098 which focused on pre-algebraic and high school level algebraic concepts.

All developmental students either tested into MA090 or into MA098. They were instructed in each course in a fixed set of concepts that progressed at the same pace for all students over the course of a 15-week semester. They were expected to get a C- or better grade in the course based on quizzes and exams in order to advance to the next level – either from MA090 to MA098 or from MA098 to one of two college level math courses (MA115 or MA121). There was no post-testing in the traditional sequence; the course grade determined advancement.

The structural changes to developmental math were as follows. In the new approach (MA 093) students continued to get some of the traditional lecture Monday, Wednesday and Friday for some portion of the 50 minute class, but the lecture component was limited to an introduction of the new (or continuing) concepts to be learned in a given day or week, and group learning time was included in the 50 minutes. A full 1 ¼ hour lab session was added to the schedule each week. And, the curriculum and instruction was linked to an online, tutorial platform with built-in assessments that tracked course content during the semester, and that was always available to students at any time of day, any day of the week. Students' grades still relied on standard exams, typically three over the course of the semester. The course grade (a C- or better) determined whether students advanced from this single developmental math course to college level math.

MA 093 initially used a pre- and post-assessment process, relying on the Accuplacer math placement tests. However, when The College moved away from the Accuplacer test as its placement instrument (to the Math SAT), fewer students had recorded initial placement scores in the college's student information system.[4] Post-testing using the Accuplacer was still used by some instructors. For this reason and because post-testing was also not used under the former two-course sequence, use of pre- and post-testing was not an option in this study.

Nonetheless, both of the technology platforms chosen (ALECS and, later, Pearson's - MyMathLab) utilize an ongoing assessment process for each set of concepts taught. A student progresses through these online, computerized tutorial systems by meeting performance criteria as measured by the computerized assessment. For each competency (module) students are initially assessed to determine their beginning understanding of specific competencies (concepts and computations) needed for the specific module and they are post-tested to determine the extent to which they demonstrate proficiency in those specific course competencies after having gone through computerized instruction.

Thus there are four ways that the new approach attempts to individualize instruction. The computerized tutorial systems is one way. Furthermore, the labs, a new feature added to the more traditional lecture component, are designed to allow students to move at their own pace. Instructors are available during lab time to instruct individual students when the computerized tutorials have not been able to get the student to understand a particular concept or set of calculations. Third, instructors continue to provide standard "office hours" and individual time to meet by appointment. Finally, having course material online means that a student can access

[4] Examination of these test scores are beyond this evaluation, although the Math Chair did indicate that most students did experience learning gain. The larger issue for him was that learning gain was not necessarily sufficient to move students to the college level given their math and learning deficits starting out.

course material at any time; students are not limited to lecture, lab or "office hours" to practice or get "help."

The lecture also included group instruction. As Hall et al (2002: 4) remind us, "strategies for flexible groupings are essential [to teaching in a differentiated classroom]." MA093 tended to group students by mixed ability. That is, students who had met a competency (or were at least more advanced) were grouped with other students who did not meet those competencies (or were less advanced). This kind of "peer-tutoring" is a recognized and important practice in the "differentiated" instructional process (Tomlinson, 1995; Hall et al, 2002).[5] However, there are others who suggest that "grouping by ability" may be a more appropriate grouping strategy.[6] Some of the findings in this study may suggest more emphasis on "grouping by ability" as the College continues to build its developmental math instructional process.

Finally, appreciating some of the findings of this study requires understanding that the shift from the two-course sequence to the one course involved integrating much of the basic skills instruction from MA 090 into the algebra instruction of MA 098. And, there are some MA 098 concepts that have been completely removed from the MA 093 course objectives (solving quadratic equations, e.g.). Some of the loss of "time on task" by consolidating 30 weeks into 15

[5] Tomlinson (1995) notes that "creating and giving task cards or assignment sheets to individuals or groups works well, as does going over an assignment with a few responsible students today so that they can share it with their groups tomorrow (32)." Furthermore, she suggests that "you can helps students learn to work collegially by suggesting that they ask a peer for clarification when they get 'stuck' (32)."

[6] Levy (2008: 163) states: "[t]here are times when grouping by ability is the most appropriate action…The teacher has taught the lesson and a small group of students need further instruction…The teacher pulls these students together for additional support…There was also a group who came into the class knowing what was taught…The teacher can pull these students together and take the lesson to the next level."

weeks is made up through integration and by including a 1 ¼ hours lab each week. In addition, the online tutorial capacity facilitates students putting as much additional time into learning and practicing difficult topics as the individual student deems he or she needs to attain mastery. But, there was some course content loss.[7]

Model, Analytical Approach and Methods

As noted above, this study takes advantage of the natural experiment that a complete cessation of one approach and the initiation of new approach offers. There are two models being tested. The first model asserts that differentiated instructional methods help students to learn better and thus helps them acquire course objectives – successful completion of a course. The second model, the core model in this study, asserts that upon completing a course, students will now be prepared to meet the demands of higher level courses and meet higher level course objectives. In terms of this study, we are asserting that differentiated instruction is more effective than traditional pedagogical approaches at getting students to developmental course completion and at getting students to complete college level courses. Both of these assertions implicitly assume that students undergoing differentiated instruction will perform better at both levels. So, it is not just course completion but also grades (grade points) that have to be examined.

To test the first model, the two developmental groups (differentiated and non-differentiated) will be compared in terms of completion rates and grade distribution at the developmental course

[7] A preliminary analysis of post-tests (Accuplacer's Elementary Algebra test) for a subset of students in MA 093 suggests that while learning gain among students in MA093 was fairly ubiquitous, those gains may not be moving all students to the college level.

level. Contingency tables and associated chi-square testing will be employed to test the significance of any differences found in course completions and in the distribution of grades. Findings here may also have implications for the findings addressing the core issue of performance at the college level.

For the latter model, the core model in this study, there are three key groups of students who will be compared: students not taking any developmental math, students taking developmental math under the original approach and students taking developmental math using the new differentiated, instructional approach. The primary dependent factors on which they all will be compared is the grade received and the rate of completion (C- or above) in 2 college level courses (MA115 - *Mathematical Ideas*, MA121 - *Elements of College Algebra*). The core question here is: do students who complete the new differentiated approach perform better in college courses than students who complete the traditional two-sequence course, both compared to students who did not take a developmental course. The main methods involved will be the generation of mean grade scores with ANOVA (and some t-test for difference of means) to determine the significance of differences found. Contingency tables showing grade distributions in the college level courses and associated chi-square testing will also be used.

Furthermore, because this is a natural experiment we cannot be sure that the students being compared are similar on factors exogenous to the model being examined. So, regression analysis will be employed to control for a range of factors that may vary between the different groups. Specific control factors include: demographic factors (gender, race/ethnicity, athlete status, other socioeconomic indicators) and measures of ability (high school GPA and standardized test scores). Logistic regression will be used to control for these exogenous factors on course completions at the developmental level, where undergoing differentiated instruction or not is the key independent variable. Hierarchical linear regression, with nesting, will be employed with respect to the second model: the first level in the hierarchical model is whether a student took and completed a developmental course, and the second level is whether the developmental courses was the differentiated instructional course.

The Data

Implementation of the newly structured developmental math course, MA093, was officially piloted in the spring semester of 2012 with 32 students. This newly structured course (MA093) was offered alongside the second semester of the originally structured developmental course sequence (MA098) which enrolled 52 students that semester. The first course of the old course sequence (MA090) was discontinued after fall semester 2011. Any student who did not pass MA090 was re-enrolled at a later date in the newly structured course. Those who did complete this level were enrolled at a later date in the final offering of MA098 in fall semester 2012.

In the following fall semester 2012, 70 students were enrolled in the new course and 71 students were enrolled in the second semester of the original course sequence (MA098). MA098 was discontinued after fall semester 2012. Since then, there have been 4 full semesters in which MA093 has been the only developmental math course offered at The College, covering the content of MA090 and MA098. Over this time there have been 149 enrollments in the newly

Table 1: Enrollments in Developmental Courses by Semester

Semester	MA 090	MA 098	MA 093
Fall 2005 to Fall 2011	562	1168	0
Spring 2012	0	52	32
Fall 2012	0	71	70
Spring 2013	0	0	30
Summer 2013	0	0	4
Fall 2013[1]	0	0	85
Spring 2014	0	0	30
TOTAL[2]	562	1291	251

1. The MA093 course in this semester was not part of Title III effort and was not structured as proposed.
2. Some students had to retake courses and/or took combinations of courses so totals are not unique students.

structured course.[8] So, there was a full break between the implementation of the new developmental course and the old two-course sequence with minor overlap of some students having been instructed in the old course sequence and in the new course. Table 1 shows the shift in tabular form.

Table 2 shows the unique student enrollments in developmental and college level math courses between fall semester 2005 and spring semester 2014. For example, the College enrolled 228 *unique* students in MA093 in that time period even though there were 251 course enrollments over that time period. Some students (23) took MA093 more than once. This is the case for all courses. So, to maintain one record per student in our dataset, only a student's last enrollment

[8] A small number of MA090 and MA098 students have been required to take MA093, having not successfully completed either MA090 or MA098.

Table 2: Number of Unique Students Taking Key Math Courses

Course	Number	Percent
No Developmental	1785	53.3%
Any Developmental	1563	46.7%
MA090	490	14.6%
MA098	1144	34.2%
MA093	228	6.8%
MA115	753	22.5%
MA121	2235	66.8%
TOTAL	3348	100.0%

Note: The numbers and percents do not reflect an official number or percent of students in any given year in developmental courses. These are the number of students across all years who had taken the specified course at least once.

record (and grade) is included for analysis.[9] Overall, course grades were collected on 3,348 unique students who had enrolled in some combination of developmental and college level courses. Nearly, 47 percent of students (N=1,563) had taken a developmental course at the college. Nearly 67 percent of all students took MA121 (N=2,235) as their college level course.

Table 3 shows the composition of college level course takers by developmental course status. Developmental course students constituted 34 percent of all college level course takers (N=936). The bulk of the developmental students in the dataset taking college level courses were students instructed in the former two-course sequence (30 percent, N=1,144). Only 4 percent of college level course takers (N=104) had been instructed in the differentiated developmental course.[10] So, there may be consequences of this small sub-sample as we progress through the analysis.

[9] The one implication is that course completions for MA093 may be understated and student performance in all courses may be overstated.

[10] Data collection for evaluating MA093 was through the spring 2014 s semester. So, many of the students who had taken MA093 in Fall 2013 (N=85) or Spring 2014 (N=30) had not yet had the opportunity to take a college level course.

Table 3: Students Taking College Level Courses by Developmental Course Status

Course	No College Level	College Level	Total Taking Specified Course	% of All College Level
Developmental Course	627	936	1563	34.4%
MA090	252	238	490	8.7%
MA098	338	806	1144	29.6%
MA093	124	104	228	3.8%
Not Developmental	0	1785	1785	65.6%
TOTAL	627	2721	3348	100.0%

The data was compiled from the College's main student information. Grades were collected for all students who enrolled in any of the three developmental courses and the two college level courses. Additional demographic information was also compiled as potential control factors (gender, race/ethnicity, family income, Pell grant receipt, athlete status) from the main student information system and to a lesser degree from the student financial aid system.[11] Students' high school grade point average and SAT scores were also compiled as indicators of students' prior academic ability.[12]

Table 4A: Developmental Course Completion (C- grade or better)

Course Type	Did Not Complete	Completed	Total	% Not Complete	Percent Complete
Traditional	366	969	1335	27.4%	72.6%
Differentiated	61	167	228	26.8%	73.2%
Total	427	1136	1563	27.3%	72.7%

Chi Square: .043 (p=.836)

[11] Developmental course students differed significantly from non-developmental course students – higher proportions in terms of being male, black or Latino, a Pell grant recipient, an athlete and first generation. These differences were greater for MA093 students, the differentiated course.

[12] Developmental students generally had lower high school GPAs and SAT scores. Students in the differentiated course (MA093) had even lower scores than developmental students generally.

Results

Developmental course completion and grade. Table 4A shows the course completion rates for developmental math students at the College by differentiated versus traditional course structure. This table measures the course completion rate by identifying all students who started in the relevant developmental course sequence (traditional in MA090 or MA098 and in MA093 for the differentiated) and determining how many of them successfully completed the course. As the table clearly shows, under both structures 27 percent of developmental students did not successfully complete their developmental course requirement. So, there is no statistically significant difference in completion rates between the traditional and differentiated students in terms of who moves out of the developmental level.

However, there are a significant number of students who may have taken and successfully completed the first course in the traditional sequence (MA090), who simply never took the second course in the sequence (MA098). When only students who took MA098 (whether as a result of being placed in MA098 or having completed MA090) the completion rate for the traditionally instructed students changes dramatically. Table 4B shows completion rates for developmental students when those MA090 students are removed. In this case, the non-

Table 4B: Developmental Course Completion (C- grade or better)
(Includes only students who eventually took MA098 in traditionl)

Course Type	Did Not Complete	Completed	Total	% Not Complete	Percent Complete
Traditional	168	969	1137	14.8%	85.2%
Differentiated	61	167	228	26.8%	73.2%
Total	427	1136	1563	27.3%	72.7%

Chi Square: 19.518 (p=.000)

completion rate for traditionally instructed students is 15 percent versus the 27 percent for differentiated students, a statistically significant difference (chi square=19.518; p=.000). So, at minimum, the move to the new MA093 differentiated approach did not improve the student completion rate and may have diminished it. But, this is not a surprising result given the compression of two courses or 30 weeks of instruction into 15 weeks for so many students.

Table 5 shows the grade distribution for students who completed the developmental courses N=969 for traditional MA098 and N=167 for differentiated MA093).[13] This table provides more detail on what "completed" means in terms of student proficiency with developmental course material. Students in the differentiated course are significantly more likely to earn a grade at the low end of the "completed the course" grade distribution. While only 7 percent of traditional students earned a C- grade (N=70 of 169), 20 percent of differentiated students earned a C- grade (N=34 of 167). The difference in those proportions was statistically significant. Ultimately, it appears that higher proportions of students passing the differentiated course passed with the

Table 5: Grade Distribution in Developmental Courses (completers)

	Number		Percent	
Grade	Tradition	Differ	Tradition	Differ
A	144	21	14.9%	12.6%
A-	100	14	10.3%	8.4%
B+	81	14	8.4%	8.4%
B	135	18	13.9%	10.8%
B-	132	21	13.6%	12.6%
C+	101	17	10.4%	10.2%
C	206	28	21.3%	16.8%
C-	70	34	7.2%	20.4%
TOTAL	969	167	100.0%	100.0%

[13] Note that the traditional students are only those who took MA098 and does not include students who did not progress to MA098 after completing the first course in the sequence, MA090.

Table 6. Effect of MA 093 on Developmental Course Completion Controlling for Key Factors (reduced equation; highly insignificant control factors not shown)

Variable	Beta	Standard Error	Significance	Exp(B)
HS GPA	1.308	.198	.000	3.700
Math SAT (/100)	.282	.145	.052	1.325
Income (/10,000)	.022	.015	.128	1.023
Took Differ	-.522	.219	.017	.593
Constant	-2.590	.749	.001	.075

Initial Percent Correct: 83.7
Predicted Percent Correct: 83.7
Did not pass percent correct: 1.2
Passed Course percent Correct: 99.8

N=1062 | Dependent: completed developmental course with a C- or better grade.

lowest grades than among students who passed the traditional sequence.[14] So, competency and proficiency may also be more limited among the differentiated students.

Because this was a natural experiment, we suspected that some of these apparent differences may result from characteristics of the two differently taught student bodies being different. Table 6 shows the reduced results of a logistic response regression with course completion (1=yes, 0=no) as the dependent variable. The results suggest that this concern, while warranted does not prove decisive in shifting the original finding regarding course completions. Taking the differentiated course appears to be negatively associated with developmental course completion even when controlling for key characteristics: the differentiated course decreases the odds of completing the developmental course requirement by over a half (.593). It is worth noting however that this equation does little increase our capacity to predict whether a student will or will not complete the developmental course requirement based on how they were taught; the equation only

[14] It is also worth noting that across the entire grade distribution differentiated students were more likely to get an F grade or withdraw from the course than students in the traditional course: 18 percent in the differentiated course versus 9 percent in the traditional course.

successfully predicted two students' non-completion (a 1.2 percent success rate in predicting non-completion). So, while there clearly is association between differentiated instruction and course non-completion, causality is fairly weak.

Table 7 shows the results of ordinary least squares regression with the final developmental course grade as the dependent variable. Note that students who withdrew from the course and did not register a grade with points associated are removed from the analysis. Nonetheless, in addition to being associated with a somewhat reduced probability of completing the developmental course, the differentiated instructional approach appears to also be associated with reduced proficiency with the developmental course material. Specifically, when controlling for exogenous factors, taking the differentiated course appears to reduce the final course grade by .20 points. This is close to moving from a B- to a C+, for example. Ultimately, the compression of two courses into one may have had the effect of trying to force students to learn too much in too short a period of time.

Table 7: Effect of MA 093 on Developmental Grades Controlling for Key Factors (reduced equation; highly insignificant control factors not shown)

Variable	Beta	Standard Error	Standard Beta	Sig
(Constant)	-.279	.271		.303
HS GPA	.684	.066	.303	.000
Math SAT (/100)	.286	.052	.159	.000
Male	-.133	.068	-.057	.052
Black	-.172	.084	-.059	.041
Latino	-.197	.109	-.050	.072
Independent	.343	.159	.059	.031
Took 093	-.204	.092	-.063	.027

R: .401

R^2: .161

N=1139 | Dependent: Grade in MA 098 or MA 093

Table 8A: Course Completion Rates (C- or Better) in Math 115 by Developmental Course Status

Developmental Status	Number	Percent	Total
No Developmental Course	400	85.7%	467
Completed MA 098	216	86.1%	251
Completed MA 093 (New)	26	74.3%	35
TOTAL	642	85.3%	753

Chi square: 3.538 (p=.171)

College level course completion and mean grade points. There are two main comparisons: the comparison of developmental students in the two different approaches to students who were not required to take a developmental course, and comparison of developmental students taking the new course to developmental students who had enrolled in the old sequence of courses. There wo college level courses are examined separately as they are two very different courses. The expectation is that students taking the new developmental course would demonstrate better completion rates and performance in terms of earned grade in college level courses than those taking developmental courses under the former approach.

Table 8A shows the college level course completion rates for students who took MA115 (Mathematical Ideas) by their developmental course taker status. As the table shows, overall 85 percent of students taking MA115 between fall semester 2005 and spring semester 2014 passed the course with a C- or better. Nearly 86 percent of students who had successfully completed MA098 (last course in the old sequence) completed MA115. In contrast, only 74 percent of

Table 8B: Course Completion Rates (C- or Better) in Math 121 by Developmental Course Status

Developmental Status	Number	Percent	Total
No Developmental Course	1421	85.2%	1668
Completed MA 098	415	80.9%	513
Completed MA 093 (New)	39	72.2%	54
TOTAL	1875	83.9%	2235

Chi square: 10.934 (p=.004)

MA093 students who had successfully completed MA 093 completed MA115. The overall differences were not statistically significant, however.

The results for students taking MA 121 (Elements of College Algebra) by developmental course status are shown in Table 8B. While 84 percent of all students taking MA 121 over the time period in question earned a C- or better in the course, only 72 percent of students who completed MA 093 earned a C- or better in MA 121. Furthermore, in contrast to the results for MA 115, students who completed MA 098 completed MA 121 at a substantially lower rate (81 percent success) than students who did not take a developmental math course (85 percent success). The differences in success between these three groups of students were statistically significant.

Table 9A and Table 9B show the mean grade points for students who took MA115 and MA121, respectively, by developmental course status. The grade points included are only those associated with grades A to F, so withdrawing students are not included. It should also be noted that MA 093 students were more likely to withdraw from MA 115 and MA 121 than MA 098 students. For example, 8 percent of MA 093 students withdrew from MA 115 versus 3 percent of students who took MA098. The differences in rates for MA115 were not statistically significant. In MA 121, 13 percent of MA 093 students withdrew from the course versus 6

Table 9A: Average Earned Gradepoints in Math 115 by Developmental Course Status

Developmental Status	Mean	Median	Std Dev	Number
No Developmental Course	2.85	3.00	1.103	360
Completed MA 093 (New)	2.30	2.00	1.134	33
Completed MA 098	2.39	2.33	0.906	250
TOTAL	2.65	2.67	1.005	643

F = 18.134 (p=.000)

Notes: t-tests show that the mean difference between MA093 and MA098 was significant at the .598 level.

percent of MA098 students. The differences for MA121 were significant. Overall, MA093 non-completers in college level courses were more likely to be withdraws than earners of Ds and Fs.

Nonetheless, Table 9A shows that there are statistically significant differences in the mean grade earned in MA 115 by developmental course status (F=18.134; p=.000). As expected, students not required to take a developmental course earned the highest mean grade of 2.85, while students taking MA093 (the differentiated course) earned a mean grade of 2.30, the lowest of all groups. The differences between mean grades for the two developmental groups were not statistically significant however, implying that the main differences in mean grades in MA 115 was between developmental students generally and non-developmental students.

Table 9B shows that there are also statistically significant differences in the mean grade earned in MA 121 by developmental course status (F=26.200; p=.000). Again, as expected, students not required to take a developmental course earned the highest mean grade of 2.80, while students taking the MA093 (the differentiated course) earned a mean grade of 2.27, the lowest of the three groups. But, while the differences between developmental students and non-developmental students were significant overall, the difference between MA093 and the MA 098 (the traditional Course) were not statistically significant. Ultimately, then then there are no statistically

Table 9B: Average Earned Gradepoints in Math 121 by Developmental Course Status

Developmental Status	Mean	Median	Std Dev	Number
No Developmental Course	2.80	3.00	1.071	1403
Completed MA 093 (New)	2.27	2.33	1.052	49
Completed MA 098 Only	2.31	2.33	0.987	496
TOTAL	2.66	2.67	1.072	1948

F = 26.200 (p=.000)
Notes: The difference between MA093 and MA098 was significant at the .766 level.

significant differences in college math performance for students who took developmental math, as long as students complete the course.

College level course grades controlling for key factors.[15] For this final analysis we use a hierarchical linear regression. The first level of analysis is the individual student enrolled in college level math and various factors that may affect any student's college level math outcomes. The second level of analysis is the developmental student with inputs into and outcomes of his or her developmental course enrollment (the grade in the developmental course, in particular).[16] Developmental students, then, may be enrolled in MA 093 or MA 098, the third level in the hierarchy. There were no additional variables associated with this third level.

Table 10A and Table 10B show the results of the hierarchical linear regressions for students in MA115 and students in MA121, respectively. The results shown are for a reduced models, meaning that a number of the control variables used are not shown.[17] Full results can be found in the appendices.[18] Note also that consistent with practice in hierarchical modeling, multiple estimations that step the different levels of factors into the equation are shown in the tables, with

[15] While we initially intended to include a logistic response regression on course completion at the college level, we decided after examining the data that examining course grade points alone would be sufficient to discern whether other factors played a role in

[16] Note that we also included HS GPA and SAT scores for only developmental students at this level to control for any possible interactions between those factors and developmental student status. This was especially important given that the college had moved from the Accuplacer as the tool that determined math placement to use of the SAT.

[17] Most factors that showed insignificant coefficients were removed from the analysis.

[18] These full specifications also show the control variables that were originally considered for the analysis and their partial effects on grade points in the courses.

consideration being given to changes in the R2 values as that occurs. We also use MA093 (the differentiated students) as the treatment in one set of estimations, while using MA098 (the traditional) in another set.

The regressions provide further insight into the relationship between developmental courses and college level course performance. First, as Table 10A (MA115) and Table 10B (MA121) show, at the first level, a student's high school GPA and his or her math SAT score has a positive effect on student performance in either of the two college level math courses. Being male has a negative effect on college level performance. Other level 1 control factors, as already noted, did not show a significant effect on grades. This first level is without respect to whether a student is a developmental student or not.

The next two levels in the hierarchy require some careful interpretation of the results and so we address the results for MA115 and MA121 separately. In MA115, having completed any developmental course is generally associated with a decrease in a student's grade. See Table 10A. Only in one specification, where the high school GPA and Math SAT for developmental students is entered into the equation, does the negative relationship prove statistically insignificant (model 3). Once the developmental course grade is entered alongside the high school GPA and math SAT, the resulting relationship between completing any developmental course and the MA 115 grade is negative and statistically significant (models 4, 5, 6, 9 and 10).

Overall, it appears that completing a developmental course captures variance in MA115 grades due to developmental students' lower overall mathematical ability. This lower ability compared

Table 10A: Hierarchical (Nested Developmental Math Course Variables) Linear Regression Results for Math 115

Variable	1	2	3	4	5	6	7	8	9	10
(Constant)	.069	.550**	.505*	.684**	.683**	.688**	.554**	.549**	.725**	.731***
	(.163)	(.289)	(.287)	(.288)	(.288)	(.288)	(.286)	(.285)	(.288)	(.288)
HS GPA	.527***	.545***	.569***	.550***	.551***	.550***	.544***	.545***	.506***	.505***
	(.073)	(.072)	(.092)	(.091)	(.091)	(.091)	(.072)	(.072)	(.073)	(.073)
Math SAT/100	.320***	.229***	.228***	.198***	.198***	.197***	.229***	.229***	.209***	.209***
	(.050)	(.054)	(.062)	(.062)	(.062)	(.062)	(.054)	(.054)	(.054)	(.054)
Male	-.275***	-.269***	-.267***	-.242***	-.242***	-.241***	-.269***	-.268***	-.248***	-.245***
	(.079)	(.078)	(.078)	(.077)	(.077)	(.077)	(.078)	(.078)	(.077)	(.077)
Completed Developmental Course		-.341***	-.039	-.622**	-.626**	-.598**	-.336***	-.310**	-1.018***	-.924***
		(.081)	(.237)	(.281)	(.285)	(.288)	(.084)	(.163)	(.227)	(.245)
High School GPA for Dev Students			-.071	-.144	-.144	-.142				
			(.123)	(.124)	(.124)	(.124)				
Math SAT/100 for Dev Students			-.034	-.049	-.049	-.044				
			(.078)	(.078)	(.078)	(.078)				
Grade in MA 098 or MA 093				.294***	.294***	.298***			.245***	.257***
				(.078)	(.079)	(.079)			(.076)	(.077)
Took and Passed 098						-.065		-.036		-.144
						(.168)		(.162)		(.164)
Took and Passed 093					.014		-.048		.017	
					(.174)		(.175)		(.175)	
R	*.429*	*.456*	*.459*	*.479*	*.479*	*.479*	*.456*	*.456*	*.471*	*.473*
R^2	*.184*	*.208*	*.211*	*.230*	*.230*	*.230*	*.201*	*.208*	*.214*	*.223*

N=579 | Dependent: earned gradepoints in Math 115

* p<=.10

**p<=.05

***p<=.01

Table 10B: Hierarchical (Nested Developmental Math Course Variables) Linear Regression Results for Math 121

Variable	1	2	3	4	5	6	7	8	9	10
(Constant)	-.327**	-.133	-.181	-.061	-.072	-.035	-.146	-.126	-.030	.001
	(.163)	(.177)	(.178)	(.177)	(.177)	(.177)	(.177)	(.177)	(.176)	(.176)
HS GPA	.545***	.539***	.528***	.519***	.518***	.515***	.538***	.534***	.512***	.504***
	(.043)	(.043)	(.048)	(.047)	(.047)	(.047)	(.043)	(.043)	(.043)	(.043)
Math SAT/100	.372***	.340***	.359***	.338***	.341***	.335***	.344***	.343***	.333***	.332***
	(.032)	(.034)	(.037)	(.037)	(.037)	(.037)	(.034)	(.034)	(.034)	(.034)
Male	-.348***	-.342***	-.344***	-.331***	-.340***	-.338***	-.351***	-.351***	-.339***	-.339***
	(.049)	(.049)	(.049)	(.048)	(.048)	(.048)	(.049)	(.049)	(.048)	(.048)
Completed Developmental Course		-.151	.234	-.665***	-.717***	-.490**	-.169***	.213	-1.175***	-.730***
		(.054)	(.164)	(.212)	(.214)	(.225)	(.055)	(.146)	(.173)	(.206)
High School GPA for Dev Students			.018	-.078	-.079	-.076				
			(.084)	(.084)	(.084)	(.084)				
Math SAT/100 for Dev Students			-.106**	-.098*	-.095*	-.066				
			(.055)	(.054)	(.054)	(.056)				
Grade in MA 098 or MA 093				.379***	.385***	.389***			.344***	.363***
				(.058)	(.058)	(.058)			(.056)	(.057)
Took and Passed 098						-.368**		-.392***		-.514***
						(.156)		(.147)		(.146)
Took and Passed 093					.342**		.308*		.371**	
					(.172)		(.174)		(.172)	
R	*.486*	*.490*	*.493*	*.511*	*.513*	*.514*	*.491*	*.493*	*.507*	*.511*
R^2	*.236*	*.240*	*.243*	*.262*	*.263*	*.264*	*.241*	*.243*	*.257*	*.261*

N=1727 | Dependent: earned gradepoints in Math 121

* p<=.10

**p<=.05

***p<=.01

to non-developmental students is neither overcome by taking a developmental math course, nor is it explained by high school GPAs or math SAT scores of developmental students: neither of those two covariates at level 2 in the hierarchy was statistically significant. Overall, being a developmental student is associated with between a .3 (a C to a C- grade, for example), and a whole point (a C to a D grade) decline in MA115 performance.

In contrast, the grade that a student earns in any developmental course is associated with a .25 to .30 increase in the grade points earned in MA115. The better a student does in the developmental math course the better he or she is likely to perform in MA 115. So, while completing a developmental math course does not put developmental students on par with non-developmental student's in MA115 performance, the better a student does in developmental math, the better they do in MA115.

Finally, at level 3 of the hierarchy, completing MA093 had no significant effect on MA 115 performance with MA 098 completion as the reference. Similarly, and as expected, completing MA 098 had no effect on MA 115 performance with MA 093 as reference. Ultimately, there is no difference in MA 115 performance of developmental students due to the type of developmental course completed.

Level 2 and level 3 results for MA 121 are more pronounced and more significant than for MA 115. Completing any developmental course tends to show a negative relationship with MA 121 grades. This result was not true for every specification: models 1, 2 and 8 do not show any statistically significant effect of taking any developmental course on Math 121 performance and models 2 and 8 both show a positive effect if any. What moves the general "developmental course completion" to a statistically significant negative relationship to MA 121 grades is the inclusion of the developmental course grade. When the developmental course grade is included in models 4,5,6, 9 and 10, it shows a significant positive effect itself, on the order of a .35 increase in grade point for MA 121 for every full grade increase in the developmental course grade, and taking a developmental course shows a consistently negative effect on the order of a .65 to 1.2 grade point decrease in MA 121. Again, the coefficient on the "completed development course" factor is measuring unknown factors related to developmental students' abilities that are not measured by the high school GPA or math SAT scores. And, this factor is negative and significant whether the level 3 factor is MA 093 or MA 098.

Furthermore, unlike for MA 115, it does matter to student performance in MA 121 whether they completed MA 093 or completed MA 098. With MA 098 as the reference, results in Table 22 (models 5, 7 and 9) show that successfully completing MA 093 consistently improves the student grade in MA 121 by .31 to .37 grade points or the difference between a C and a C+ grade in MA 121. Conversely, MA 098 has a negative relationship to students' MA 121 grade. So, when controlling for a range of covariates unlike the earlier findings from the descriptive statistics, MA 093 does improve student performance in MA 121 significantly.

Discussion and Considerations of the Differentiated Model Adopted

From a formative standpoint, The College has successfully transformed its developmental course approach from a traditional "lecture and test" approach to one built on a differentiated learning approach. Most elements of this new approach have been successfully implemented as proposed. Preliminary (although incomplete) data, as noted above, suggests that most students undergoing instruction in MA093 are experiencing significant levels of learning gain. But this study does raise some concerns about whether students who are allowed to progress to the college level based on the grade in the differentiated learning course are truly college-ready.

The move from the MA090 and MA098 developmental course sequence to MA 093 as an integrated developmental course using a differentiated learning approach has not been a complete boon to student outcomes to the extent that those outcomes are being measured by course grades. MA093 has not appreciably increased developmental course completion rates as hoped. It is also true that completion rates have not plunged with the shift to MA093. Given the compression of two courses, the first of which (MA090) served a substantially less prepared student than the second (MA098), into one course, it could be argued that completion rates should have plummeted. But they did not.

Furthermore, MA093 has not been a boon for college level math completion for developmental students. Specifically, the withdrawal rates for MA093 students in college level courses are somewhat higher for completers of MA093 than for completers under the former regimen. But, the drop in completion is only on the order of 3 or 4 developmental students per 100. And, when MA093 students do not withdraw, they do tend to perform as well in terms of their college level grade, particularly in MA121, as under the old course sequence, all other things being equal. In fact, what the hierarchical regressions suggest is that at every level of developmental course grade, students coming out of the differentiated course (MA093) do somewhat better than students from the traditional course at the same developmental grade level.

Ultimately, given that the move to MA093 neither dramatically improves student outcomes nor diminishes them, another criterion of success should at least be considered. That criterion is simple. By moving from two courses to one course, with no dramatic decline in developmental course completion at the developmental or college levels, the college decreased the amount of time students spend at the developmental level, potentially shortening their time to graduation and saving the students money.

That does not suggest that there are no areas for improving the impact on learning gain and preparedness. In that light here are some suggestions that reflect on both the findings in this study and the literature on differentiated learning.

1) Given higher rate of F grades in MA093 and assuming that some portion of the higher rate of withdraws in MA093 are to avert a pending F, more emphasis should be placed on direct instruction to struggling students. For example, more classroom grouping by similar ability rather than mixed ability may allow instructors more time even during lecture (as opposed to just lab) to focus on students having the most difficulty, and *who may also be the least self-directed*.
2) Use of grades to determine whether a student moves to the college level from MA093 should be complemented for at least some students (say, those with less than a B grade) with an exit exam.

3) For students who do not pass the exit exam and who have low grades, the college may want to consider an alternative grade given that developmental courses count toward the student's overall cumulative GPA.

4) Consistent with recommendation (3) the college should emphasize to developmental students that MA093 is intended for them to accelerate their learning at the developmental level, if they can and want to do so, but that not completing should not be seen as a "failure" as long as substantial learning gain is taking place. Some students are starting from very far behind and simply need more than a semester to get to college level (presupposed under the old two sequence course regime). One idea would be to keep two grades for each student: one grade would be the standard grade measuring student performance on exams, quizzes, etc.; while a second grade would focus more on individual student learning gain. A weighted combination of these two grades would constitute a semester course grade.

References

Hall, Tracey et al. 2002. "Differentiated Instruction and Implications for UDL Implementation: Effective Classroom Practices Report." National center for Accessing the General Curriculum in Cooperative Agreement with the Department of Education, Office of Special Education.

Huebner, Tracy A. 2010. "What Research Says About Meeting Students Where They Are/Differentiated Learning." Meeting Students Where They Are." Educational Leadership; 67 (5). Association for Supervision and Curriculum Development.

Levy, Holli M. 2008. "Meeting the Needs of All Students through Differentiated Instruction: Helping every Child Reach and Exceed Standards." The Clearing House: A Journal of Educational Strategies, Issues and Ideas; 81(4: 161-164).

Raudenbush, Stephen W. 1993. "Hierarchical Linear Models and Experimental Design" in Applied Analysis of Variance in Behavioral Science. (p. 459-496). Marcel Dekker; New York.

Tomlinson, Carol Ann. 1995. How to Differentiate Instruction in Mixed-Ability Classrooms. Association for Supervision and Curriculum Development (ASCD); Alexandria, VA.

Woltman, Heather et al. 2012. "An Introduction to Hierarchical Linear Modeling." Tutorials in Quantitative Methods for Psychology. 8(1: 52-69).

Subban, Pearl. 2006. "Differentiated Instruction: A Research Basis." International Education Journal. 7(7), 935-947. Shannon Research Press.

Washburn, Carleton W. 1953. "Adjusting the Program to the Child." Educational Leadership. Association for Supervision and Curriculum Development.

Dosch, Mary and Margaret Zidon. 2014. "'The Course Fit Us:' Differentiated Instruction in the College Classroom." International Journal of Teaching and Learning in Higher Education.

APPENDICES

Appendix A: Full Linear Regression Models on Developmental Course Outcomes

Dependent: Grade Point in Developmental Course

Variable	**Beta**	**Standard Error**	**Standard Beta**	**Sig**
(Constant)	-.324	.321		.313
HSGPA	.698	.069	.308	.000
MathSAT100	.360	.059	.197	.000
VerbalSAT100	-.074	.054	-.044	.174
Football	.161	.105	.051	.126
Male	-.220	.080	-.094	.006
Black	-.193	.092	-.067	.036
Latino	-.217	.114	-.056	.057
Income	.007	.008	.043	.403
EFC	.000	.000	-.032	.525
FirstGen	-.001	.068	.000	.994
Independent	.354	.162	.063	.029
Pell	.035	.087	.015	.689
Took093	-.200	.095	-.062	.037
R	*.415*			
R^2	*.172*			

Appendix B: Grade Distribution in Math 115 by Developmental Course Sequence

	Number					Percentage				
Grade	No Develop Course	Took MA 093	Took MA 098	Took MA 090 Only	Took MA 090 and MA 098	No Develop Course	Took MA 093	Took MA 098	Took MA 090 Only	Took MA 090 and MA 098
A	67	5	10	1	0	15.0%	13.9%	5.2%	7.7%	0.0%
A-	62	1	14	2	2	13.9%	2.8%	7.2%	15.4%	3.1%
B+	59	3	17	0	4	13.2%	8.3%	8.8%	0.0%	6.3%
B	63	3	22	2	14	14.1%	8.3%	11.3%	15.4%	21.9%
B-	37	1	25	1	7	8.3%	2.8%	12.9%	7.7%	10.9%
C+	43	1	31	0	9	9.6%	2.8%	16.0%	0.0%	14.1%
C	40	10	30	3	17	9.0%	27.8%	15.5%	23.1%	26.6%
C-	12	2	13	2	7	2.7%	5.6%	6.7%	15.4%	10.9%
D+	3	0	2	0	2	0.7%	0.0%	1.0%	0.0%	3.1%
D	5	3	7	0	1	1.1%	8.3%	3.6%	0.0%	1.6%
D-	8	3	4	0	0	1.8%	8.3%	2.1%	0.0%	0.0%
F	17	1	11	0	1	3.8%	2.8%	5.7%	0.0%	1.6%
W	30	3	8	2	0	6.7%	8.3%	4.1%	15.4%	0.0%
TOTAL	**446**	**36**	**194**	**13**	**64**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Note: the difference in withdrawal rate between MA 093 and MA098 only was not statistically significant (p=.277) while the difference in withdrawal rate between MA 093 and the combination of MA090/MA098 was statistically significant at the .019 level.

Appendix C: Grade Distribution in Math 121 by Developmental Course Sequence

	Number					Percentage				
Grade	No Develop Course	Took MA 093	Took MA 098	Took MA 090 Only	Took MA 090 and MA 098	No Develop Course	Took MA 093	Took MA 098	Took MA 090 Only	Took MA 090 and MA 098
A	351	2	27	1	3	21.5%	3.6%	6.6%	5.6%	2.6%
A-	157	4	27	1	3	9.6%	7.1%	6.6%	5.6%	2.6%
B+	160	1	27	3	4	9.8%	1.8%	6.6%	16.7%	3.4%
B	223	10	49	1	15	13.6%	17.9%	11.9%	5.6%	12.9%
B-	152	6	42	3	12	9.3%	10.7%	10.2%	16.7%	10.3%
C+	118	5	51	3	22	7.2%	8.9%	12.4%	16.7%	19.0%
C	152	8	66	4	25	9.3%	14.3%	16.1%	22.2%	21.6%
C-	82	4	42	0	9	5.0%	7.1%	10.2%	0.0%	7.8%
D+	19	1	4	0	2	1.2%	1.8%	1.0%	0.0%	1.7%
D+	45	2	13	0	6	2.8%	3.6%	3.2%	0.0%	5.2%
D-	32	2	13	2	5	2.0%	3.6%	3.2%	11.1%	4.3%
F	64	4	23	0	6	3.9%	7.1%	5.6%	0.0%	5.2%
W	79	7	27	0	4	4.8%	12.5%	6.6%	0.0%	3.4%
TOTAL	**1634**	**56**	**411**	**18**	**116**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Note: the difference in withdrawal rate between MA 093 and MA098 only approached statistical significance at p=.109, while the difference in withdrawal rate between MA 093 and the combination of MA090/MA098 was statistically significant at the .023 level.

Appendix D: Full Linear Regressions on College Level Gradepoints with All Level 1 Control Variables Entered

MA 115

	Treatment = MA 093				Treatment = MA 098			
Variable	**Beta**	**Standard Error**	**Standard Beta**	**Sig**	**Beta**	**Standard Error**	**Standard Beta**	**Sig**
(Constant)	.412	.361		.255	.405	.361		.263
HS GPA	.547	.078	.286	.000	.548	.078	.287	.000
Math SAT	.188	.066	.140	.005	.188	.067	.140	.005
Verbal SAT	.065	.064	.047	.305	.066	.064	.048	.300
Male	-.314	.090	-.155	.001	-.313	.090	-.155	.001
Black	-.163	.120	-.059	.174	-.169	.120	-.061	.157
Latino	-.169	.137	-.050	.219	-.172	.137	-.051	.211
Football	.037	.139	.012	.791	.039	.139	.012	.778
Income	.000	.000	.029	.596	.000	.000	.028	.603
First Generation	-.078	.081	-.039	.340	-.076	.081	-.038	.348
Independent	.277	.182	.061	.129	.284	.182	.063	.119
Pell	.102	.108	.051	.344	.099	.108	.050	.358
Completed Developmental Course	-.308	.087	-.153	.000	-.328	.175	-.163	.060
Took and Passed 098					.010	.174	.005	.953
Took and Passed 093	-.107	.190	-.023	.573				
R	*.479*				*.478*			
R^2	*.229*				*.229*			

MA 121

	Treatment = MA 093				Treatment = MA 098			
Variable	**Beta**	**Standard Error**	**Standard Beta**	**Sig**	**Beta**	**Standard Error**	**Standard Beta**	**Sig**
(Constant)	.189	.215		.379	.204	.215		.343
HS GPA	.531	.045	.289	.000	.527	.045	.287	.000
Math SAT	.393	.042	.264	.000	.391	.042	.263	.000
Verbal SAT	-.105	.037	-.077	.005	-.104	.037	-.076	.005
Male	-.362	.058	-.169	.000	-.362	.058	-.169	.000
Black	-.029	.073	-.010	.693	-.031	.073	-.010	.668
Latino	-.069	.084	-.019	.411	-.074	.084	-.020	.382
Football	-.012	.080	-.004	.882	-.009	.080	-.003	.914
Income	.000	.000	-.005	.866	.000	.000	-.004	.880
First Generation	-.058	.048	-.028	.229	-.057	.048	-.028	.234
Independent	-.040	.135	-.007	.767	-.042	.135	-.007	.754
Pell	-.016	.058	-.008	.778	-.016	.058	-.007	.788
Completed Developmental Course	-.157	.056	-.067	.005	.149	.148	.064	.316
Took and Passed 098					-.311	.149	-.131	.037
Took and Passed 093	.279	.172	.037	.106				
R	*.493*				*.494*			
R^2	*.244*				*.244*			

DOES SIZE MATTER?: TEXT BOX SIZE IN ONLINE SURVEYS[1,2]

Jennifer A. May, Research Analyst

Lauren M. Conoscenti, Senior Research Analyst

Office of Institutional Research & Evaluation

Tufts University

Abstract

It is well-established that a survey's visual elements can cue respondents and thus influence responses, yet most of this research has focused on closed-ended items. What do respondents think when they encounter open-ended items such as, "What is your favorite part of Dining Services?" How do associated visual cues influence respondents, if at all? Does a larger text box prompt a longer response? Is such a response more thoughtful or is it merely filling space? This project explored how and to what extent text box size influences various qualities of responses to open-ended items with an eye toward recommendations for designing open-ended items in IR surveys.

Introduction

Online surveys are perhaps the most common way of gathering data from students, alumni, and other university constituents. Most surveys contain both closed-ended and open-ended questions, but the survey research literature tends to focus on designing closed-ended questions. This study sought to understand how and to what extent survey design techniques can influence various qualities of responses to open-ended survey items in order to make recommendations for open-ended items in institutional research (IR) surveys.

[1] This study was funded in part by an NEAIR Research Grant.
[2] The authors would like to thank Dana Silverberg and Christian Testa for their invaluable assistance with data preparation and other aspects of this project.

Review of the Literature

Importance of Open-Ended Survey Items

Open-ended items serve special purposes in IR and other survey research. They allow participants to respond to items in their own words, providing rich textual support and embellishment for closed-ended item results. They are also ideal when there is scant knowledge or established literature on a subject. Allowing respondents who select “other” in response to a closed-ended question to specify what “other” means to them can create a better experience for respondents and provide researchers with more accurate information (Dillman, Smyth, & Christian, 2009; Goynea, 2005). Data from open-ended items can even satisfy constituents’ desire for qualitative data when limited resources do not permit focus groups or other qualitative data collection strategies.

Visual Cues in Online Surveys

In addition to being mindful of best practices for all types of surveys (such as clear writing and manageable length), researchers deploying online surveys must also consider the impact of visual cues on data quality. In fact, this may be especially important in online surveys given the increased ease, and thus tendency, for survey designers to incorporate visual elements (Dillman, Christian, & Smyth, 2009; Tourangeau, Couper, & Conrad, 2004). Past studies have found that the layout of a matrix question (e.g. Couper, Tourangeau, Conrad, & Zhang, 2013), mobile-optimized survey formats (e.g. Stapleton, 2013), and the use of pictures to supplement question text (e.g. Couper, Conrad, & Tourangeau, 2007; Toepoel & Couper, 2011) can affect the quantity and quality of information gathered. It has also become common practice to use radio buttons to denote a closed-ended item requiring only one response and check boxes to denote a closed-ended item allowing multiple responses. Since text boxes are another type of visual cue, it follows that their features might influence the quantity and quality of responses (e.g., Behr, Bandilla, Kaczmirek, & Braun, 2014; Christian, Dillman, & Smyth, 2007).

Past Studies of Text Boxes

The limited amount of prior research on text boxes has focused on open-ended items requesting numerical responses, and has generally found that various design elements, including size, influence responses. For example, Christian, Dillman, and Smyth (2007) demonstrated that respondents prompted with "Month" and "Year" beneath text boxes were less likely to provide this information in a two-digit month/four-digit year format than those prompted with "MM" and "YYYY". Similarly, Couper, Kennedy, Conrad, and Tourangeau (2011) showed that respondents reported currency more precisely when prompted with a text box bracketed by a "$" symbol and ".00". Dillman, Smyth, and Christian (2009) summarized several studies on the influence of text box size, concluding that larger boxes nearly always encouraged longer answers, whether appropriate for the question or not; for instance, in questions requiring a numerical input, such as number of hours per week studying, a larger box was more likely to lead to a range, such as 2.5–5, which the researcher then had to recode, estimate, or omit.

Yet many open-ended items require a narrative or descriptive answer, and these question types have been less well-studied. Christian and Dillman (2004) and Dillman, Smyth, and Christian (2009) found that text box size can cue respondents to the expected length of the response; specifically, survey participants wrote more words and discussed more topics in larger text boxes than they did for smaller text boxes. Smyth, Dillman, Christian, and McBride (2009) replicated this finding for late survey responders, but not for early responders. However, Behr, Bandilla, Kaczmirek, and Braun (2014) found that, in a study of a specific cognitive probe, or a follow-up item probing why a respondent chose a particular answer to a closed-ended item, larger text boxes resulted in unwanted or unusable information. Couper, Kennedy, Conrad, and Tourangeau (2011) also noted the importance of text box size, stating, “The designer needs to decide on the size of the text box, thereby encouraging shorter or longer responses,” but failed to provide more specific guidance (p. 68). None of these studies considered the possibility of varying text box size within the same survey when different items necessitated different length and depth of responses.

Clearly, manipulating the text box size has an effect, although what effect it may have remains murky, and perhaps not as well-explored as institutional researchers and others conducting survey research might like. Additionally, conflicting studies have introduced the possibility that different types of open-ended items, such as true narrative items versus follow-up probes, may yield different responses.

What Does the IR Literature Say?

Evidence-based recommendations for formatting open-ended items are especially scant in the IR literature. Many publications exist to guide institutional researchers through the process of designing and administering surveys specifically in a higher education setting (e.g., Chekis-Gold, Loescher, Shepard-Rabadam, & Carroll, 2006; Suskie, 1996; Porter, 2004; Umbach, 2005). Unfortunately, these offer only a tertiary discussion, if any, of open-ended items. Often, the only recommendation is to simply avoid such items.

Institutional researchers also struggle with declining response rates due in part to “survey fatigue” (e.g., Porter & Whitcomb, 2005). Dillman, Smyth, and Christian (2009) found that open-ended items are especially likely to contribute to survey fatigue and non-response; however, as discussed, carefully chosen and well-written open-ended items fulfill purposes that closed-ended items cannot. This suggests the need for guidance on strategies to reduce the burden of responding to open-ended items while ensuring that they meet research goals.

Research Questions

This study explored the following research questions:

1. Does the size of a text box influence quantitative measures of open-ended item data quality, including survey completion, item response rate, and length of responses?
2. Does the size of a text box influence qualitative measures of open-ended item data quality, notably response content and tone or valence of responses?

Methods

Institutional Context

Tufts University is a private research institution that has four campuses (three in Massachusetts and one in France) and grants bachelor's, graduate, and professional degrees. Tufts attracts academically talented first-time, full-time freshmen. Each year, over 1,300 students graduate with bachelor's degrees and the institution has a consistent four-year graduation rate of 85% ± 2% (Freeman, Sharkness, & Terkla, 2015).

Experimental Design

This study employed an experimental design manipulating text box size in two undergraduate surveys, the Orientation Survey and the Dining Services Satisfaction Survey, that Tufts' Office of Institutional Research and Evaluation (OIRE) administered in Fall 2014. For each survey, a random half of participants received a version with large text boxes (600 pixels wide by 90 pixels high) for all narrative and probe open-ended items while the other half received small text boxes (400 pixels wide by 30 pixels high) for the same items. Note that both sizes of text box allowed respondents an unlimited number of characters.

Both surveys featured a mix of narrative and probe open-ended item types. Narrative items appeared to all respondents and were not explicitly related to any closed-ended items; common narrative survey items include questions about strengths and weaknesses of a particular campus program or the, “Any additional comments?” question that appears at the end of many surveys. Follow-up probes only appeared to respondents who selected certain responses to preceding closed-ended items and asked respondents to explain those closed-ended responses. In these surveys, such items typically appeared when a respondent selected a negative response (e.g., “very dissatisfied,” “disagree”) to the corresponding closed-ended question.

Measures and Analysis

This study considered five measures of open-ended item data quality, three quantitative (response length, survey completion rate, and item response rates) and two qualitative (response content or whether or not a respondent explained an answer, and tone or valence). We used Statistical Package for Social Sciences (SPSS) Version 22 to conduct independent samples t-tests comparing respondents receiving large and small text boxes on response length, and to conduct chi-square tests of independence on survey completion rate and item response rate. We used Linguistic Inquiry and Word Count (LIWC) software to analyze qualitative dimensions, and chi-square tests in SPSS to compare prevalence of different qualitative characteristics.

Surveys

Orientation Survey. In Fall 2014, this survey contained a total of 25 open-ended items, eight narrative questions and 17 follow-up probes. (See Table 1 for a complete list of open-ended items.) The survey yielded an overall response rate of 49.6% (N = 331); of the respondents, 48.9% had received large text boxes and 51.1% had received small text boxes. This difference of 2.2 percentage points was not significant ($\chi^2 = 0.15$, $p = 0.70$).

Table 1

Orientation Survey Open-Ended Items[3]

Narrative Items
Which social activity during Orientation did you like best and why?
Please provide any comments or feedback you have that might be useful for future Orientation Leaders, ACE Fellows, or Resident Assistants.
Please provide any comments or feedback you have that might be useful for future Pre-Orientation Leaders.
What were the highlights of your Pre-Orientation experience?
What would you change about your Pre-Orientation experience?
Of the Orientation programs that you attended, which would you like follow-up on during your first year at Tufts?
Was there anything you expected to be covered during Orientation that was not covered? If so, please explain.
Do you have any additional comments about Orientation?
Follow-Up Probes
If this session [Introducing the Departments and Programs, Academic Essentials, Academic Integrity Workshops, Faculty Forums, Speak About It, Many Stories, One Community, Common Reading Book, Operation Awareness] was not useful, please explain why.
If you were dissatisfied or very dissatisfied with your individual advising session with your academic advisor, please explain why.
If you were dissatisfied with the registration process, please explain why.
If you were dissatisfied with the Orientation Office/Orientation Hotline, please explain why.
Why didn't you use *Student Connection* Tufts' First-Year Student website?
If you or your family did not find *Student Connection* useful, please explain why.
If you were dissatisfied with the Student Services Desk, please explain why.
How, if at all, was your Pre-Orientation experience different than what was advertised?
If you did not apply to and/or participate in a Pre-Orientation program, why not?
If yes, how did not participating in a Pre-Orientation program affect your experience?

[3] Complete instrument is available upon request.

Dining Services Satisfaction Survey. This survey featured 16 open-ended items, six narrative questions and 10 follow-up probes, and yielded a response rate of 39.9% (N = 1,019). (See Table 2 for a complete list of open-ended items.) Among respondents, 49.1% had received large text boxes and 50.9% had received small text boxes, and, not surprisingly, this small difference was not significant ($\chi^2 = 0.48$, $p = 0.51$).

Table 2

Dining Services Satisfaction Survey Open-Ended Items[4]

Narrative Items
What other foods or beverages would you like to have available at [Carmichael/Dewick-MacPhie, Hodgdon, Brown & Brew]?
What do you think of the lunchtime burrito bar?
What is your favorite thing about Tufts Dining?
Please use the space below to provide any additional comments you have about on-campus dining.
Follow-Up Probes
If you were dissatisfied with your experience at [Carmichael, Dewick-MacPhie, Hodgdon, Pax et Lox, Hotung Café, Brown & Brew, Tower Café], please indicate why below.
You indicated that you have not eaten at or purchased food from some of the Tufts Dining locations. Please tell us why you do not visit those locations—and what might encourage you to visit in the future.
Is there anything else that we can do to better accommodate your dietary needs?
Is there any additional information you would like to see on Tufts Dining Social Media or other Social Media we should be using to connect with you?

[4] Complete instrument is available upon request.

Results

Quantitative Measures of Data Quality

Respondents receiving larger text boxes wrote longer responses than respondents receiving smaller text boxes. Of the 26 items examined, larger text boxes yielded longer responses in 23 cases. Differences were statistically significant for five items:

- Which social activity during Orientation did you like best and why?
- What other foods or beverages would you like to have available at Hodgdon?
- What other foods or beverages would you like to have available at Brown & Brew?
- If you were dissatisfied with your experience at Dewick-MacPhie, please indicate why below.
- If you were dissatisfied with your experience at Hodgdon, please indicate why below.

That these five yielded statistically significant results is not surprising given the salience of the social aspect of Orientation and of these particular dining locations at Tufts and related higher response rates to these items.

There were no significant differences in survey completion rates or item response rates between the group receiving large and the group receiving small text boxes. There were no apparent trends in the types or nature of questions for which large or small text boxes tended to yield higher item response rates.

See Tables 3-6 for complete results.

Table 3

Item Response Rates (RR) and Mean Word Counts by Text Box Size, Orientation Survey Narrative Items

	Large Text Boxes				Small Text Boxes				
			Word Count				Word Count		Mean
Item	N	RR	Mean	SD	N	RR	Mean	SD	Difference
Which social activity during Orientation did you like best and why?	106	65.4%	15.24	13.72	105	62.1%	11.01	9.09	4.23*
Please provide any comments or feedback you have that might be useful for future Orientation Leaders, ACE Fellows, or Resident Assistants.	35	21.6%	21.11	16.12	27	16.0%	17.15	16.11	3.96
Please provide any comments or feedback you have that might be useful for future Pre-Orientation Leaders.	26	31.0%	15.92	16.25	19	22.6%	15.37	11.40	0.55
What were the highlights of your Pre-Orientation experience?	61	72.6%	11.57	12.04	66	78.6%	8.33	9.83	3.24
What would you change about your Pre-Orientation experience?	52	61.9%	8.19	9.42	56	66.7%	6.27	8.20	1.92
Of the Orientation programs that you attended, which would you like follow-up on during your first year at Tufts?	76	46.9%	5.72	7.58	75	44.4%	3.40	3.59	2.32
Was there anything you expected to be covered during Orientation that was not covered? If so, please explain.	61	37.7%	6.75	8.87	60	35.5%	4.50	6.20	2.25
Do you have any additional comments about Orientation?	40	24.7%	7.37	12.63	51	30.2%	8.12	11.93	-0.75

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
Mean Difference = Large - Small

Table 4

Response Rates and Mean Word Counts by Text Box Size, Orientation Survey Follow-Up Probes

Item	Large Text Boxes				Small Text Boxes				Mean Difference
			Word Count				Word Count		
	N	RR	Mean	SD	N	RR	Mean	SD	
If this session [Introducing the Departments and Programs] was not useful, please explain why.	5	50.0%	--	--	3	50.0	--	--	--
If this session [Academic Essentials] was not useful, please explain why.	5	41.7%	--	--	7	46.7	--	--	--
If this session [Academic Integrity Workshops] was not useful, please explain why.	20	64.5%	--	--	8	44.4	--	--	--
If this session [Faculty Forums] was not useful, please explain why.	1	33.3%	--	--	2	28.6	--	--	--
If this session [Speak About It] was not useful, please explain why.	8	66.7%	--	--	2	28.6	--	--	--
If this session [Many Stories, One Community] was not useful, please explain why.	6	40.0%	--	--	9	45.0	--	--	--
If this session [Common Reading Book] was not useful, please explain why.	20	62.5%	10.05	8.35	2	60.6	9.60	6.17	0.45
If this session [Operation Awareness] was not useful, please explain why.	4	57.1%	--	--	2	33.3	--	--	--
If you were dissatisfied or very dissatisfied with your individual advising session with your academic advisor, please explain why.	10	100.0%	--	--	8	80.0	--	--	--

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
Mean Difference = Large - Small

Table 4, cont.

Response Rates (RR) and Mean Word Counts by Text Box Size, Orientation Survey Follow-Up Probes, cont.

	Large Text Boxes				Small Text Boxes				
			Word Count				Word Count		Mean
Item	N	RR	Mean	SD	N	RR	Mean	SD	Difference
If you were dissatisfied with the registration process, please explain why.	21	91.3%	--	--	18	100.0%	--	--	--
If you were dissatisfied with the Orientation Office/Orientation Hotline, please explain why.	1	100.0%	--	--	2	100.0%	--	--	--
Why didn't you use Student Connection Tufts' First-Year Student website?	7	63.6%	--	--	11	64.7%	--	--	--
If you or your family did not find Student Connection useful, please explain why.	1	50.0%	--	--	2	100.0%	--	--	--
If you were dissatisfied with the Student Services Desk, please explain why.	1	100.0%	--	--	1	50.0%	--	--	--
How, if at all, was your Pre-Orientation experience different than what was advertised?	7	53.8%	--	--	4	100.0%	--	--	--
If you did not apply to and/or participate in a Pre-Orientation program, why not?	49	30.2%	9.27	7.04	53	31.4%	7.23	4.68	0.19
If yes, how did not participating in a Pre-Orientation program affect your experience?	31	96.9%	14.65	8.78	39	95.1%	13.69	12.03	0.95

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
Mean Difference = Large - Small

Table 5

Response Rates (RR) and Mean Word Counts by Text Box Size, Dining Survey Narrative Items

Item	Large Text Boxes N	Large Text Boxes RR	Large Text Boxes Word Count Mean	Large Text Boxes Word Count SD	Small Text Boxes N	Small Text Boxes RR	Small Text Boxes Word Count Mean	Small Text Boxes Word Count SD	Mean Difference
What other foods or beverages would you like to have available at Carmichael/Dewick-MacPhie?	218	51.2%	12.39	24.81	249	53.9%	6.66	6.78	5.73
What other foods or beverages would you like to have available at Hodgdon?	123	53.2%	9.50	10.70	135	54.4%	6.45	6.23	3.04*
What other foods or beverages would you like to have available at Brown & Brew?	26	31.7%	10.08	16.47	32	29.4%	7.04	8.35	3.04
What do you think of the lunchtime burrito bar?	163	70.6%	5.93	6.27	174	70.2%	5.39	5.68	0.54
What is your favorite thing about Tufts Dining?	248	50.0%	9.37	9.81	285	54.5%	6.95	7.31	2.42*
Please use the space below to provide any additional comments you have about on-campus dining.	126	25.4%	25.68	34.65	159	30.4%	20.83	29.15	4.85

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
Mean Difference = Large - Small

Table 6

Response Rates (RR) and Mean Word Counts by Text Box Size, Dining Survey Follow-Up Probes

Item	Large Text Boxes				Small Text Boxes				Mean Difference
			Word Count				Word Count		
	N	RR	Mean	SD	N	RR	Mean	SD	
If you were dissatisfied with your experience at Carmichael, please indicate why below.	80	51.2%	15.30	22.80	249	53.9%	13.71	17.49	1.59
If you were dissatisfied with your experience at Dewick-MacPhie, please indicate why below.	104	28.7%	20.79	31.64	125	32.1%	11.29	12.90	9.50*
If you were dissatisfied with your experience at Hodgdon, please indicate why below.	52	22.5%	22.53	26.61	60	24.2%	14.34	15.40	8.20*
If you were dissatisfied with your experience at Pax et Lox, please indicate why below.	24	27.3%	19.69	17.41	26	32.1%	13.80	11.97	5.89
If you were dissatisfied with your experience at Hotung Café, please indicate why below.	24	18.8%	30.14	46.47	28	20.0%	21.05	25.02	9.09
If you were dissatisfied with your experience at Brown & Brew, please indicate why below.	15	18.3%	13.62	8.69	19	17.4%	19.63	21.05	-6.01
If you were dissatisfied with your experience at Tower Café, please indicate why below.	28	20.3%	20.58	20.52	28	19.2%	16.71	23.32	3.87
You indicated that you have not eaten at or purchased food from some of the Tufts Dining locations. Please tell us why you do not visit those locations—and what might encourage you to visit in the future.	6	100.0%	--	--	2	100.0%	--	--	--

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
Mean Difference = Large - Small

Table 6, cont.

Response Rates (RR) and Mean Word Counts by Text Box Size, Dining Survey Follow-Up Probes, cont.

	Large Text Boxes				Small Text Boxes				
			Word Count				Word Count		Mean
Item	N	RR	Mean	SD	N	RR	Mean	SD	Difference^
Is there anything else that we can do to better accommodate your dietary needs?	16	51.6%	19.05	21.57	21	56.8%	16.81	18.30	2.23
Is there any additional information you would like to see on Tufts Dining Social Media or other Social Media we should be using to connect with you?	15	48.4%	6.36	7.77	16	43.2%	6.34	6.91	0.02

* Indicates difference is statistically significant, $p < .05$.
-- Indicates too few responses to support significance testing.
^Mean Difference = Large - Small

Qualitative Measures of Data Quality

Response content. This analysis compared respondents who explained their answers to open-ended questions (in other words, answered "what" and "why") to those who did not explain their answers (in other words, only answered "what"). For two of the three items that lent themselves to this type of comparison, those receiving large text boxes were significantly more likely to explain "why." For example, 68.6% of large text box respondents answered the "why" part of the question, "Which social activity during Orientation did you like best and why?" compared to only 31.4% of small text box respondents ($\chi^2 = 4.65$, $p < .05$).

This trend persisted even when the question did not prompt respondents to explain their answers. The question, "What is your favorite thing about Tufts Dining?" did not ask respondents to explain their answers, but 74.3% of large text box respondents did so compared to only 58.7% of small text box respondents, a statistically significant difference ($\chi^2 = 16.34$, $p < .05$).

Tone or valence of responses. Large text boxes tended to yield a greater proportion of responses with negative valences for questions that did not imply a particular tone. For example, "Please use the space below to provide any additional comments you have about on-campus dining" yielded significantly more negative comments among those provided with large text boxes ($\chi^2 = 7.94$, $p < .05$). Specifically, 64.5% of large text box respondents provided a comment with a negative valence compared to only 40.7% of small text box respondents. This pattern held for three additional items, "Do you have any additional comments about Orientation?" (61.5% of large text box respondents provided purely negative responses compared to 47.6% of small text box respondents), "Please provide any comments or feedback you have that might be useful for future Orientation Leaders, ACE Fellows, or Resident Assistants" (68.8% vs. 60.9%), and, "Please provide any comments or feedback you have that might be useful for future Pre-Orientation Leaders" (20.8% vs. 10.5%), though the differences were not statistically significant. The lack of statistical significance is likely due to the overall small numbers of respondents who had negative comments on these items. Note that, for each question, the vast majority of responses were either clearly positive or clearly negative; only a minority could be considered neutral or nonresponsive and thus those few responses were excluded from the analysis.

See Table 7 for complete results.

Table 7

Response Tone/Valence Distribution (%) by Text Box Size

	Large Text Boxes			Small Text Boxes			
Item	N	Positive	Negative	N	Positive	Negative	χ^2
Please use the space below to provide any additional comments you have about on-campus dining.	62	35.5%	64.5%	81	59.3%	40.7%	7.94*
Please provide any comments or feedback you have that might be useful for future Pre-Orientation Leaders.	24	79.2%	20.8%	19	89.5%	10.5%	.83[1]
Please provide any comments or feedback you have that might be useful for future Orientation Leaders, ACE Fellows, or Resident Assistants.	32	31.3%	68.8%	23	39.1%	60.9%	.37
Do you have any additional comments about Orientation?	13	38.5%	61.5%	21	52.4%	47.6%	.62

*Statistically significant, $p < .05$.
For all items, analysis excluded neutral responses.
1. Cell sizes did not meet assumptions.

Discussion

Does Size Matter?

As expected, the size of the text box affected several features of responses to open-ended survey items. Quantitatively, size affected response length such that respondents wrote significantly more words per item and wrote more overall across all survey items when they saw large text boxes instead of small ones. This finding is consistent with literature demonstrating that respondents are more likely to write more even if the question does not demand it as in the case of providing four-digit years instead of two-digit years when given a larger space for the digits (Christian, Dillman, and Smyth, 2007).

In this study, longer responses generally provided additional, new information, often in the form of “why,” and were not necessarily just “fillers.” Yet whether or not writing more words results in more meaningful or useful responses depends on the intent of the survey. For surveyors who want additional information about why particular services or experiences were dissatisfying or ineffective, a longer response may be helpful. For researchers who only want a simple list, extra words may complicate interpretation with unnecessary data.

Perhaps more significantly, respondents receiving larger text boxes differed in the nature of the answers they provided in two important ways. On narrative items, respondents receiving large text boxes were significantly more likely to address “why” or otherwise explain their answers when they received large text boxes whether or not the question prompted them to explain their answers. Using a large text box seemed to cue respondents to provide more than a one-word or one-phrase response even if the question itself did not cue them to do so.

Respondents were also significantly more likely to write negative responses when they received large instead of small text boxes. Although the reason for this is unclear, one possible explanation might be that the large size reinforced respondents' propensity to use the survey as a sounding board or "rant." Established survey research literature suggests that those who had very positive or very negative experiences are most likely to respond to a survey; perhaps the very negative group viewed the large box as an opportunity to express their vehement dissatisfaction (e.g., Dillman, Smyth, & Christian, 2009).

However, it is also important to highlight the domains in which text box size did *not* matter: item response rates or survey completion. This is a particularly important finding because it suggests that text box size did not contribute to survey fatigue or nonresponse bias, two perennial problems in survey research.

Recommendations for Survey Research

In general, the results of this experiment suggest that survey researchers should design text boxes at an individual item level such that the text box is proportional to the nature of the response the researcher seeks. Elaboration might provide useful detail for some items, in which case a survey researcher should provide respondents with a large or even oversized text box to reinforce this message. For a different item on the same survey, a one-word response or simple list might suffice; researchers should size boxes accordingly to avoid misleading respondents or collecting unnecessary information. Since text box size did not significantly impact whether a respondent answered an open-ended item or answered subsequent closed- or open-ended items, size decisions can be made at the item rather than the survey level.

For institutional researchers, the challenge may be determining or helping a campus constituent determine what type of response would be most helpful when asking a particular open-ended question. Does the constituent need only the name of the social activity or food? Perhaps a single-line text box is best to limit the amount of unnecessary information collected.

Does the constituent want to know why students had a negative Orientation experience? Not only should "why" be explicitly included in the question, but the text box should be relatively large to further cue students that a longer answer is expected.

Limitations of this Study

This study has several limitations. The sample included only undergraduate students at an elite research institution, and thus may not be generalizable to other student populations at the university or to other campuses. Instead, this study is best viewed as a tool for increasing awareness of the potential impact of text box size on open-ended responses and a starting point for survey design. Other institutional researchers might conduct similar studies to confirm, qualify, or refute these findings for their own student populations.

Additionally, although the surveys in this study yielded relatively high response rates, nonresponse bias may have impacted results (Croninger & Douglas, 2005; Sax, Gilmartin, & Bryant, 2003; Tschepikow, 2012; Umbach, 2005). This is especially likely given the prevalence of nonresponse to open-ended items compared to closed-ended items. Finally, non-completion bias may have interfered with the study's results since, consistent with best practices for survey research, most open-ended narrative items in these surveys appeared toward the end of the survey (Dillman, Smyth, & Christian, 2009).

Opportunities for Future Research

Planned next steps in this study include another set of experiments designed to confirm, refute, or qualify the findings presented here. Specifically, subsequent experiments will evaluate whether the finding that a larger text box prompts respondents to discuss "why" or otherwise explain their answers holds using other experimental manipulations. Additionally, we will explore whether this study's findings hold when text box sizes vary randomly over the course of the survey.

Other researchers might consider replicating these experiments with different surveys, populations, and institutional contexts. Such experiments would aid in determining whether this study's results are generalizable to contexts other than the one employed here. With increasing proportions of respondents completing surveys on mobile devices and early findings that mobile respondents are less likely to respond to open-ended items—and type less when they do—it is vital that future research also considers text box size in the context of mobile devices (Buskirk & Andrus, 2012; Lambert, 2015).

Conclusion

Clearly, like other visual features of online surveys, text box size can systematically affect responses. This suggests the need for survey researchers to carefully consider the goal of asking each open-ended question and likely ways in which the data will be used in order to make a thoughtful decision when including a text box. Perhaps this quotation from Richard Linklater summarizes it best: "Whatever story you want to tell, tell it at the right size."

References

Behr, D., Bandilla, W., Kaczmirek, L., & Braun, M. (2014). Cognitive probes in web surveys: On the effect of different text box size and probing exposure on response quality. *Social Science Computer Review, 32(4)*, 524-533. doi: 10.1177/0894439313485203

Buskirk, T.D., & Andrus, C. (2012). Smart surveys for smart phones: Exploring various approaches for conducting online mobile surveys via smartphones. *Survey Practice, 5*(1). http://surveypractice.org/index.php/SurveyPractice/article/download/63/pdf.

Christian, L. M., & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly 68(1)*, 57–80. doi: 10.1093/poq/nfh004

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping respondents get it right the first time: The influence of words, symbols, and graphics in web surveys. *Public Opinion Quarterly, 71(1)*, 113-125. doi:10.1093/poq/nfl039

Chekis-Gold, R., Loescher, R., Sehpard-Rabadam, E., Carroll, B. (2006, Winter). Essential steps for web surveys: A guide to designing, administering, and utilizing web surveys for university decision-making. *Association for Institutional Research Professional File, 102*. Retrieved from http://www.airweb.org/EducationAndEvents/Publications /ProfessionalFiles/Pages/default.aspx

Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual context effects in web surveys. *Public Opinion Quarterly, 71*(4), 623–634. doi: 10.1093/poq/nfm044

Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. *Journal of Official Statistics, 27*(1), 65–85. Retrieved from http://www.degruyter.com/view/j/jos

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review, 31(3),* 322-345. doi: 10.1177/0894439312469865

Croninger, R.G., & Douglas, K.M. (2005). Missing data and institutional research. In P.D. Umbach (Ed.), Survey research: Emerging issues. *New Directions for Institutional Research, 127*, 33-49.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Freeman, L. R., Sharkness, J., & Terkla, D. G. (2015). *Tufts University Fact Book 2014-2015*. Medford, MA: Tufts University, Office of Institutional Research & Evaluation.

Goynea, R.M. (2005). Self-reported data in institutional research: Review and recommendations. In P.D. Umbach (Ed.), Survey research: Emerging issues. *New Directions for Institutional Research, 127*, 73-89.

Lambert, A.D., & Miller, A.L. (2015). Living with smartphones: Does completion device affect survey responses? *Research in Higher Education, 56*, 166-177. doi: 10.1007/s11162-014-9354-7

Porter, S.R. (2004). Raising response rates: What works? In S.R. Porter (Ed.), Overcoming survey research problems. *New Directions for Institutional Research, 121*, 5-21.

Porter, S.R., & Whitcomb, M.E. (2005). Non-response in student surveys: The role of demographics, engagement, and personality. *Research in Higher Education, 46*(2), 127-152. doi: 10.1007/s11162-004-1597-2

Sax, L.J., Gilmartin, S.K., & Bryant, A.N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44*(4), 409-432. doi: 0361-0365/03/0800-0409/0

Smyth, J.D., Dillman, D.A., Christian, L.M., & McBride, M. (2009). Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly, 73*(2), 325-337. doi: 10.1093/poq/nfp029

Stapleton, C. (2013). The smartphone way to collect survey data. *Survey Practice, 6*(2). Retrieved from http://www.surveypractice.org/index.php/SurveyPractice/article/view/75

Suskie, L. (1996). *Questionnaire survey research: What works*. (2nd ed.) . Tallahassee, FL: Association for Institutional Research.

Toepoel, V., & Couper, M. P. (2011). Can verbal instructions counteract visual context effects in web surveys? *Public Opinion Quarterly, 75(1),* 1-18. doi: 10.1093/poq/nfq044

Tourangeau, R., Couper, M.P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68*(3), 368-393. doi: 10.1093/poq/nfh035

Tschepikow, W.K. Why don't our students respond? Understanding declining participation in survey research among college students. *Journal of Student Affairs Research and Practice, 49*(4), 447-462. doi: 10.1515/jsarp-2012-6333

Umbach, P.D. (2005). Getting back to the basics of survey research. In P.D. Umbach (Ed.), Survey research: Emerging issues. *New Directions for Institutional Research, 127*, 91-100.

IR Reports

The NEAIR Best IR/Practitioner Report Award was instituted to recognize and promote quality reports that are presented at the annual NEAIR conference, but are not necessarily in the scholarly paper format necessary for submission to the Best Paper/Best First Paper Awards.

Submissions for the Best IR/Practitioner Report Award do not typically fit the scholarly research paper model. Instead, they are applied projects incorporating innovative research and solutions to specific IR problems. These papers are driven by a research question, emphasizing novel solutions that would benefit the typical IR office. While this work may be grounded in an understanding of the IR literature and best practices, a literature review is not necessary.

Similarly, the work may not necessarily involve a formal research protocol or advanced statistical analysis. However, all analyses undertaken should be appropriate to the problem or research question

Examples of projects appropriate for this category include enrollment reports, results of a student survey, a market analysis, a research project undertaken to support a campus decision, etc. This list is not intended to be all-inclusive, and we welcome creative approaches and unique projects.

Using Rasch Analysis to Review the Quality of Rating Scales

Carol Van Zile-Tamsen

The State University of New York at Buffalo

Abstract

Rasch Analysis is a very useful tool to aid in reviewing the psychometric quality of rating scales and informing scale improvements. This paper presents a brief overview of the appropriate Rasch models for use with measures containing polytomous items, software options for conducting Rasch Analysis, and appropriate sample sizes for psychometric studies using Rasch models based on the stakes of the decisions to be made based on the measures. In addition, this paper outlines a five-step process for using Rasch Analysis to review the psychometric properties of a rating scale. The Partial Credit Model and Andrich Rating Scale Model will be described in terms of the pyschometric information (i.e., reliability, validity, and item difficulty) and diagnostic indices generated. Further, these principles will be illustrated through the example of authentic data generated from a university-wide course student evaluation of teaching.

Introduction

Rasch Analysis, based on Item Response Theory (IRT; Embretson & Reise, 2000), is a very useful tool for providing information about the psychometric properties of measures. The original Rasch model was developed for use with dichotomously scored items (i.e., those that are marked as either correct or incorrect), and is based on the early work of Thurstone and Guttman (Osterlind, 2009). Unlike in classical test theory, where the standard error of measurement is assumed to be equivalent across all test takers, in IRT, measurement error is assumed to vary across individuals. Estimates of the latent trait being measured are based on both person and item characteristics, and both person ability and item difficulty are measured on the same scale (logits). Thus, we can use analyses based on IRT to help us determine if item difficulties are appropriate to person ability levels on the latent trait. By more appropriately matching item difficulties to person abilities, IRT allows us to develop measures with greater score reliability using fewer test items.

The Andrich Rating Scale Model (RSM; Andrich, 1978) is a variation of the traditional Rasch model used for polytomous data (e.g., likert-type items). As with all Rasch models, information is provided about item difficulty, person ability, and reliability. In the case of a non-achievement measure, difficulty refers to how much of the latent trait the individual must possess before they positively endorse an item. Reliability information is provided for both item measurement and person measurement in the form of separation indices and reliability indices. Item separation and reliability estimates indicate the degree to which the item estimates are expected to remain stable in a new sample. In general, an item separation index greater than 3.0 coupled with reliability greater than 0.90 is an indication that the hierarchical structure of items according to level of latent trait will be stable in a new sample (Bond & Fox, 2012). The criteria for stability of item difficulty are most likely to be achieved with large sample sizes and items that have a wide range of levels of the latent trait (Linacre, 2014a).

Person reliability indices reflect the degree to which people in new samples can be classified along the latent trait being measured, and stability of classification is found when the person separation index is greater than 2.0 and the reliability estimate is greater than 0.80 (Linacre, 2014a). The person-level estimates indicate the level of generalizability of the measurement to new samples.

The Andrich Rating Scale Model provides detailed information about the behavior of individual scale options for rating scales. When using this model to estimate latent scores, diagnostic indices are generated that allow us to examine how each option is operating in terms of complete and precise measurement of the latent construct in question. The indices of interest include category frequencies and average measures, infit and outfit mean squares, and threshold calibrations. By using the RSM, we can determine if we have sufficient or too few rating scale options for the level of precision in measurement required.

The Partial Credit Model (PCM; Wright & Masters, 1982) was developed to allow for the compilation of items on different scales into an overall latent score using linking items that are on a common scale. Through the use of linking items and the PCM, it is possible to ensure that several different versions of a rating scale are measuring a latent trait in an equivalent fashion (Bond & Fox, 2012). This paper will illustrate how the PCM can be used in conjunction with the RSM to compare several rating scales to each other to select the most appropriate version. Five steps (see Figure 1) will be described for the review of the psychometric quality of rating scales according to objective criteria. However, this same review can be performed on a single rating scale without the steps involving PCM – this parallel process will be highlighted throughout the discussion below.

[INSERT FIGURE 1 ABOUT HERE.]

Step 1: Identification of a A Scaling Question or Issue

For most users of Rasch Analysis, the question or issue that brings them to Rasch involves the quality of an established rating scale. The purpose of the analysis will be to establish reliability and ensure that is indeed measuring the construct with precision. However, a very valuable use of the Andrich Rating Scale model is the information it provide about how the options in a likert-type scale are functioning. Through an examination of category diagnostic indices, a great deal of information can be gleaned about the functioning of the scale itself in providing adequate measurement (Bond & Fox, 2012).

In the illustrative example, Rasch Analysis was used to identify the most appropriate number of rating scale options for a student evaluation instrument of teaching. The Faculty Senate of a large research university in the northeast had recently adopted a common online course evaluation form to be used across all courses at the institution. The questions (shown in Table 1) were based on the educational quality factors identified by Marsh (1983, 1984, 1987) in his work with the Student Evaluation of Educational Quality (SEEQ) instrument. For each of the 11 questions, the neutral option was excluded as a strategy to encourage students to express either a positive or negative opinion. The 4-point scale included the following options: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree. Further, all 11 items were required, with no mechanism for students to opt out if they truly had no basis for forming an opinion.

[INSERT TABLE 1 ABOUT HERE.]

This approach to scale development achieved its intended purpose of maximizing collected data for those students who completed the evaluation. Every student making it to the end had a complete set of data for these 11 items because there was no way to skip questions. However, many instructors and students alike expressed concerns about the fact that students were forced to respond to all items, even when they could not form an opinion, and many students expressed feeling pressured to complete their evaluations. From a measurement perspective, the extent of bias in responses was unknown. How many students were simply selecting any response to proceed with and complete the evaluation, and were students tending to mark on the positive side or on the negative side or both?

The Faculty Senate was unwilling to revise the scale since it had taken such a long time to come to university-wide consensus on the items, the scale, and the platform. As a result, an experiment was conducted to compare several versions of the scale to determine if a more appropriate measurement scale could be identified in the hopes that they would be convinced by research to change the scale. The existing scale was compared to versions of the scale that included a midpoint and/or an opt-out option to determine if these variations impact the latent measurement of course and instructional effectiveness and to identify the most appropriate rating scale. The variations of the rating scale compared were:

Version 1: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree

Version 2: 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, 5 = strongly agree

Version 3: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree, 5 = don't know/not applicable

Version 4: 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, 5 = strongly agree, 6 = don't know/not applicable

In this example, the psychometric issue involved finding the most appropriate measurement scale among several variations. However, in most instances, the issue will involve only an examination of the psychometric properties of a single rating scale.

Step 2: Appropriate Sample Sizes and Collection of Data

To ensure sufficient responses across all of the scale options for each item, a sufficiently large sample is required. Linacre (2014b) has prepared guidelines for appropriate sample size and suggests a minimum of 10 respondents for each scale point to achieve adequate statistical power. In the present study, at the very minimum, 60 respondents are required for each of the four conditions, or 240 total respondents. A sample of this size allows for item calibration precisions within +/- ½ logit ($\alpha < .05$). As the decisions based on the measurement results become more serious, the desired measurement precision will be greater. However, the greatest number of respondents indicated by Linacre, even at the most serious levels of decision making, is 500.

With regard to the course evaluation example, oversampling was needed due to traditionally poor response rates. At this institution, course evaluation response rates per class range from 30-40%. For the present study, large undergraduate sections of seated courses (150 or more students enrolled) offered in the fall 2014 semester were identified for the course pool. This pool was further narrowed to include only sections with a single instructor. The final pool consisted of 36 courses. Ten of these instructors consented to participate (27.8%), and two of them volunteered additional course sections, resulting in a final sample of 1,271 completed course evaluations. The total student response rate across all four conditions was 43.4%.

Once instructors consented to participate, the student enrollments for the identified section(s) were randomly assigned to one of four conditions based on the version of the rating scale the students would see on the evaluation form. In addition to the 11 common course evaluation items, all study participants received five additional linking items that used the 6-point scale (midpoint and opt out), selected from the University Course Evaluation Item Bank (Purdue University Center for Instructional Excellence, 2014):

1. Relationships among course topics are clearly explained.
2. My instructor makes good use of examples and illustrations.
3. My instructor indicates relationship of course content to recent developments.
4. My instructor effectively blends facts with theory.
5. Difficult concepts are explained in a helpful way.

These items are used to link all versions of the scales together so that overall ratings of course and teacher effectiveness can be estimated on the same scale using the partial credit model and compared, regardless of condition (Step 3; Linacre, 2014a). This additional step is not required for projects where only the psychometric properties of a single scale are examined.

In sum, students in each class section involved in the data collection process randomly received one of four variations of the course evaluation rating scale, but all received the five common linking items rated on the 6-point scale. All student responses were completely anonymous and instructor identifiers were stripped from the data before data analysis began.

Step 3: Using the Partial Credit Model to Ensure Comparability of Measures

In Step 3, latent course and instructional effectiveness scores for each respondent were estimated using the Rasch PCM (Linacre, 2014a), a step that is not required when one is examining the psychometric properties of a single version of a rating scale. In the course evaluation example, five additional linking items used the rating scale with all possible options, allowing Winsteps to calibrate all responses regardless of rating scale condition and to estimate measures of course and instructional effectiveness for all respondents across all conditions. Each respondent's estimated latent Course Effectiveness and Teaching Effectiveness score was then saved to a data file that could be exported to SPSS.

Two, full factorial analysis of variance (ANOVA) models were estimated, one with Course Effectiveness as the dependent variable, and one with Instructional Effectiveness as the dependent variable. In both analyses, course section was used as a control factor to allow for the fact that different kinds of courses and different instructors will likely have different course ratings. Results of the two ANOVA's are shown in Tables 2 and 3. With regard to measures of Course Effectiveness, controlling for section effects, the main effect of rating scale condition was not statistically significant ($F_{(3,1242)} = 0.69$), indicating that the format of the scale does not impact measures of course effectiveness.

[INSERT TABLE 2 ABOUT HERE.]

[INSERT TABLE 3 ABOUT HERE.]

The result was similar for the effect of rating scale format on measures of Instructional Effectiveness ($F_{(3,1242)} = 0.85$), indicating that, when instructor differences are taken into account, measures of instructional effectiveness are equivalent across the four versions of the scale. In this analysis, however, both the main effect for course section and the interaction effect were significant. One section, in particular, seemed to have a much different pattern of measures across the four scales. A review of the raw data revealed that ratings for this course, regardless of scale, were much lower than ratings for the other courses, at least one standard deviation lower in most cases. This outlier appears to be the cause of the significant interaction, since this effect becomes non-significant when this section is removed from the analysis ($F_{(30, 1188)} = 1.36$).

This finding suggests that overall latent measures of instructional effectiveness are also consistent across all scales once outliers (one course section) are excluded. Based on the findings of these two ANOVA's, we can proceed to Step 4.

Step 4: Examining Rating Scale Diagnostics and Reliability Indices

In Step 4, Winsteps (Linacre, 2014a) is used to run the Andrich RSM (Andrich, 1978; Bond & Fox, 2012) and generate rating scale diagnostics and reliability indices. This step is relevant for examinations of psychometric quality. For examinations of a single rating scale, this analysis will be run just one time. Options that are considered opt out options, such as "don't know" or "not applicable," are coded as missing values in these analyses. Procedures and resulting fit indices outlined by Bond and Fox (2012) are used to analyze the measurement precision of the rating scale. These include item separation and reliability and person separation and reliability. Category diagnostics are examined to determine the appropriateness of the number of response options for each scale, including category frequencies and average measures, infit and outfit mean squares, and threshold calibrations. Probability curves, showing the likelihood of responses for each response option, are generated to provide a visual analysis of the appropriateness of each option.

[INSERT TABLE 4 ABOUT HERE.]

The item separation and reliability estimates and person separation and reliability estimates for the course evaluation example are shown in Table 4. As mentioned above, an item separation index greater than 3.0 coupled with reliability greater than 0.90 is an indication that the hierarchical structure of items according to difficulty level will be stable in a new sample. With regard to Course Effectiveness, the item reliability indices do not achieve these criteria for stability of item difficulty across samples. The separation and reliability estimates are extremely consistent across Conditions 1, 3, and 4, with Condition 2 having the lowest item separation and reliability estimates. This lack of item stability is likely due to the fact that all of the items on this measure are very closely clustered together in terms of difficulty. In contrast, item reliability estimates for the Instructional Effectiveness measure do achieve these criteria for item stability across all four versions of the scale.

Three of the four versions of the Course Effectiveness measure have adequate person reliability (i.e., person separation greater than 2.0 and reliability greater than 0.80), Conditions 1, 2, and 4. For Instructional Effectiveness, only Conditions 1 and 2 have adequate person reliability. The low values could indicate that the sample did not contain a wide enough variation in opinions about course and instructional effectiveness or additional items are needed for this measure.

Tables 5 and 6 include the category diagnostic indices for each category within each condition for the two measures. In terms of category frequencies, each category should have at least 10 responses, and average measures should increase monotonically from the lowest rating point to the highest rating point. Infit and outfit mean squares should be less than 2.0; values higher than this suggest that the category is not contributing to the measurement of the latent trait and, in fact, may be working to diminish precision. Finally, with regard to thresholds, each threshold, or step up the scale, should be at least 1.4 logits greater than the last to show appropriate distinction between categories. However, intervals of more than 5 logits indicate that there is a gap in the measurement of the trait.

[INSERT TABLE 5 ABOUT HERE.]

As Table 5 shows, for Course Effectiveness, each version of the scale meets the criteria for category frequency and monotonicity of average measures. The lowest category frequencies are for the ‘don’t know/not applicable’ option in Conditions 3 and 4, but each of these still exceeds the minimum criterion of 10. Further, all of the infit and outfit mean squares are less than 2.0. Thus, the thresholds appear to be the index of most value for determining the appropriateness of each of the four scales. For Condition 1, the threshold distance between points 1 and 2 and 2 and 3 fall within the appropriate range of widths (2.99 and 1.43, respectively), but the distance between 3 and 4 (6.11) suggests that another option would be appropriate between these two. This pattern is similar for Condition 3, but with a slightly smaller distance between 3 and 4 (5.14). For Condition 2, all threshold distances are of appropriate size except for the distance between 4 and 5, which is slightly larger than desirable (5.1). In Condition 4, all threshold distances meet the criterion. Probability curves showing category frequencies and thresholds are shown in Figure 1. These curves illustrate the data shown in Table 5: for every version of the scale, respondents are most likely to be grouped in the top two categories. In the versions used in Conditions 2 and 4, the neutral midpoint appears to have a minimal role, but threshold distances between the last two options are smallest when the midpoint is included.

[INSERT FIGURE 2 ABOUT HERE.]

[INSERT TABLE 6 ABOUT HERE.]

Each version of the Instructional Effectiveness scale also meets the criteria for monotonicity of average measures, but, in Condition 4, the ‘don’t know/not applicable’ option does not achieve the minimum of 10 respondents (see Table 6). As with the Course Effectiveness measure, all of the infit and outfit mean squares are less than 2.0. In Conditions 1 through 3, the threshold distances between the last two scale points exceed the maximum of 5.0 (5.32, 5.43, and 5.37, respectively). In Condition 4, the threshold distances fall within desirable levels. The probability curves showing category frequencies and thresholds for Instructional Effectiveness are shown in Figure 2. The patterns are very similar to those for Course

Effectiveness, with most respondents gravitating toward the top two options. For Instructional Effectiveness, however, the only version with appropriate threshold distance between the top two options is Condition 4, which includes both a neutral midpoint and ‘don’t know/not applicable.’

[INSERT FIGURE 3 ABOUT HERE.]

Based on the evidence provided, the Faculty Senate did adopt the proposed rating scale options for the university-wide course evaluation. The Course Effectiveness measure now uses a five-point scale with a neutral midpoint, and the Instructional Effectiveness measures uses a six-point scale with both a neutral midpoint and ‘don’t know/not applicable.’ The overall university response rate is staying steady at about 42.0%, but the students are much more comfortable with the new response options, and “complaints” from students during the course evaluation administration period about the scale are now non-existent.

Step 5: Review of Item Difficulties

The final step in the psychometric analysis is an examination of the individual items. Through an examination of the item separation index and the probability curves, we can begin to get a sense of the range of difficulty levels of items included in the measure. In an appropriately designed measure, respondents at all levels of the latent trait will be matched to items that assess their level of that trait, and we should see a full range of item difficulties. Item separation indices that do not meet the criteria of 3.0 suggest that the difficulty levels of the items may be mismatched with respondents. Additional evidence of inappropriate item difficulties may be seen in the probability curves. Taking the course evaluation results as an example, none of the versions of the scale met the item separation criteria of 3.0. In Figure 2, regardless of the version of the scale, respondents from all levels of perceptions of course effectiveness, from very low levels of perceptions that the course is effective to very high levels, selected the “agree” option, which was the most common option in each of the scales.

Wright maps illustrate how the difficulty of items, measured in logits, is matched to the overall level of the latent trait in each respondent, also measured in logits (Bond & Fox, 2012). Wright maps for the maintained scale versions of the course effectiveness and instructional effectiveness measure are shown in Figures 4 and 5. For course effectiveness, the majority of students are clustered at +6.0 or +7.0 logits, at the highest levels of perceptions of course effectiveness. Item difficulties, however, never exceed 1.0 logits, and all of the items are grouped together at the same levels of difficulty. The instructional effectiveness items have a better range of difficulties (as shown in Figure 5), but still do not exceed a 1.0 logits, and again, students are clustered at the highest levels of perceptions of instructional effectiveness. These results suggest that items with a greater range of difficulty levels are needed for both measures. With the existing measures, it takes very low levels of perceived course and instructional effectiveness to rate courses and instructors highly (positive bias; Darby, 2008).

[INSERT FIGURE 4 ABOUT HERE.]

[INSERT FIGURE 5 ABOUT HERE.]

Conclusions

This paper and the course evaluation example illustrate how Rasch Analysis can be used to successfully (and empirically) review the psychometric properties and quality of rating scales. Through the use of a systematic process to collect and analyze the data and compare results against specific, predetermined criteria, we can make conclusions about the quality of our rating scales and our items. The Andrich Rating Scale Model provides dioagnostic indicators for each response option that indicate if each is working optimally to precisely measure the construct.

The Partial Credit Model is helpful to compare different versions of scales to determine which is providing the most precise and most reliable measurement of the construct. We can use information from these two forms of analyses to work iteratively to review, revise, and refine our measures until they achieve the level of measurement precision needed for our decision-making purposes.

References

Andrich, D. (1978). Rating formulation for ordered response category. Psychometrika, 43(4), 561-573.

Bond, T. G., Fox, C. M. (2012). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Embretson, S. A., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahweh, NJ: Lawrence Erlbaum Associates.

Darby, J. A. (2008). Course evaluations: A tendency to respond 'favourably' on scales? *Quality Assurance in Education, 16*, 7-18.

Linacre, J. M. (2014a). A user's guide to Winsteps Ministeps Rasch-Model computer programs. Chicago, IL: Author.

Linacre, J. M. (2014b). Sample size and item calibration (or person measure) stability. Institute for Objective Measurement. Retrieved 10/7/2014 from http://www.rasch.org/rmt/rmt74m.htm

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology, 75*, 150-166.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-754.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11* (3), whole issue.

Osterlind, S. J. (2009). Modern measurement: Theory, principles, and applications of mental appraisal. Upper Saddle River, NJ: Pearson.

Purdue University Center for Instructional Excellence (2014). PICES item catalog. Retrieved 8/26/2014 from http://www.purdue.edu/cie/data/pices.html

Wright, B. D., & Masters, G. D. (1982). *Rating scale analysis*. Chicago, IL: Mesa Press.

Table 1: Core Course Evaluation Items

Course Effectiveness (Cronbach's α = 0.95)
1. The course was well organized.
2. The course was intellectually challenging and stimulating.
3. The work load in the course was reasonable and appropriate.
4. Methods of evaluating student work were fair and appropriate.
5. The course content (assignments, readings, lectures, etc.) helped me meet the learning expectations set forth by the instructor.
6. Overall, this was an excellent course.
Instructional Effectiveness (Cronbach's α = 0.96)
1. The instructor clearly presented what students should learn (the expected learning outcomes) for the course.
2. The instructor was enthusiastic about teaching the course.
3. The instructor made students feel welcome in seeking help/advice in or outside of class.
4. The instructor presented material clearly.
5. Overall, this was an excellent instructor.

Table 2: ANOVA Results – Course Effectiveness Measure

Source	Sum of Squares	df	Mean Square	F
Intercept	3082.54	1	3082.54	469.70***
Section	1423.78	11	129.43	19.72***
Course	13.53	3	4.51	.69
Interaction	276.23	33	8.37	1.28
Error	8151.08	1242	6.57	
Total	17375.44	1290		

Table Notes: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 3: ANOVA Results –Instructional Effectiveness Measure

Source	Sum of Squares	df	Mean Square	F
Intercept	4374.51	1	4374.51	586.74***
Section	2217.34	11	201.58	27.04***
Condition	19.02	3	6.34	.85
Interaction	404.65	33	12.26	1.65*
Error	9259.87	1242	7.46	
Total	22936.38	1290		

Table Notes: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4: Rasch Reliability Indicators for Course and Instructional Effectiveness Measures by Condition

Condition	N	Course Effectiveness				Instructional Effectiveness			
		Item		Person		Item		Person	
		Separation	Reliability	Separation	Reliability	Separation	Reliability	Separation	Reliability
1	332	2.13	0.82	2.26	0.84	4.13	0.94	2.16	0.82
2	313	1.60	0.72	2.20	0.83	3.62	0.93	2.15	0.82
3	294	2.18	0.83	1.94	0.79	3.41	0.92	1.90	0.78
4	343	2.30	0.84	2.10	0.81	4.07	0.94	1.91	0.79

Table 5: Course Effectiveness Ratings -- Response Category Fit Statistics by Condition

Condition	Category	Observed Count	Average Measure	Infit Mean Square	Outfit Mean Square	Threshhold
1	1 (SD)	98	-3.01	0.91	0.98	None
	2 (D)	160	-1.12	0.99	0.71	-2.99
	3 (A)	976	1.67	0.90	0.95	-1.56
	4 (SA)	758	4.60	1.05	0.77	4.55
2	1 (SD)	70	-1.83	0.81	1.05	None
	2 (D)	116	-0.90	1.04	0.98	-2.25
	3 (U)	128	0.02	0.85	0.78	-0.57
	4 (A)	827	1.53	0.91	0.94	-1.14
	5 (SA)	737	3.88	1.24	0.88	3.96
3[1]	1 (SD)	52	-2.38	0.76	0.62	None
	2 (D)	121	-0.38	1.10	0.95	-2.66
	3 (A)	707	1.68	0.97	1.00	-1.24
	4 (SA)	833	4.35	1.00	0.88	3.90
	5 (DK/NA)	27				
4[1]	1 (SD)	81	-1.82	0.79	0.81	None
	2 (D)	154	-0.80	1.09	1.19	-2.18
	3 (U)	155	0.04	0.83	0.79	-0.40
	4 (A)	799	1.62	0.82	0.87	-0.83
	5 (SA)	831	3.40	1.30	1.00	3.41
	6 (DK/NA)	14				

Legend: SD = Strongly Disagree, D = Disagree, U = Undecided, A = Agree, SA = Strongly Agree, DK/NA = Don't Know/Not Applicable

Notes: [1] The RSM was run with the don't know/not applicable option coded as missing data.

Table 6: Instructional Effectiveness Ratings -- Response Category Fit Statistics by Condition

Condition	Category	Observed Count	Average Measure	Infit Mean Square	Outfit Mean Square	Threshhold
1	1 (SD)	57	-4.09	0.84	0.83	None
	2 (D)	109	-1.69	0.91	0.74	-3.72
	3 (A)	659	2.24	0.92	0.96	1.60
	4 (SA)	830	5.84	1.07	0.88	5.32
2	1 (SD)	56	-2.94	0.69	0.70	None
	2 (D)	62	-1.30	1.20	1.42	-2.39
	3 (U)	115	-0.28	0.77	0.67	-1.50
	4 (A)	554	1.98	0.97	0.93	-0.77
	5 (SA)	778	5.01	1.21	0.81	4.66
3[1]	1 (SD)	56	-3.09	0.94	0.87	None
	2 (D)	87	-0.95	1.08	0.90	-2.87
	3 (A)	430	1.89	1.01	1.01	-1.25
	4 (SA)	849	4.93	0.92	0.83	4.12
	5 (DK/NA)	13				
4[1]	1 (SD)	68	-1.91	1.02	1.20	None
	2 (D)	91	-1.19	0.81	0.70	-2.13
	3 (U)	109	0.06	0.84	0.86	-0.76
	4 (A)	538	1.75	0.92	1.02	-0.79
	5 (SA)	890	3.83	1.27	0.96	3.67
	6 (DK/NA)	9				

Legend: SD = Strongly Disagree, D = Disagree, U = Undecided, A = Agree, SA = Strongly Agree, DK/NA = Don't Know/Not Applicable

Notes: [1] The RSM was run with the not applicable/don't know option coded as missing data.

Figure Captions

Figure 1: Steps in Using Rasch Analysis to Review Psychometric Properties of Rating Scales

Figure 2: Course Effectiveness -- Probability Curves of Response Categories by Condition

Figure 3: Instructional Effectiveness -- Probability Curves of Response Categories by Condition

Figure 4: Course Effectiveness – Wright Map Showing Item Difficulty vs. Person Ability

Figure 5: Instructional Effectiveness – Wright Map Showing Item Difficulty vs. Person Ability

Figure 1: Steps in Using Rasch Analysis to Review Psychometric Properties of Rating Scales

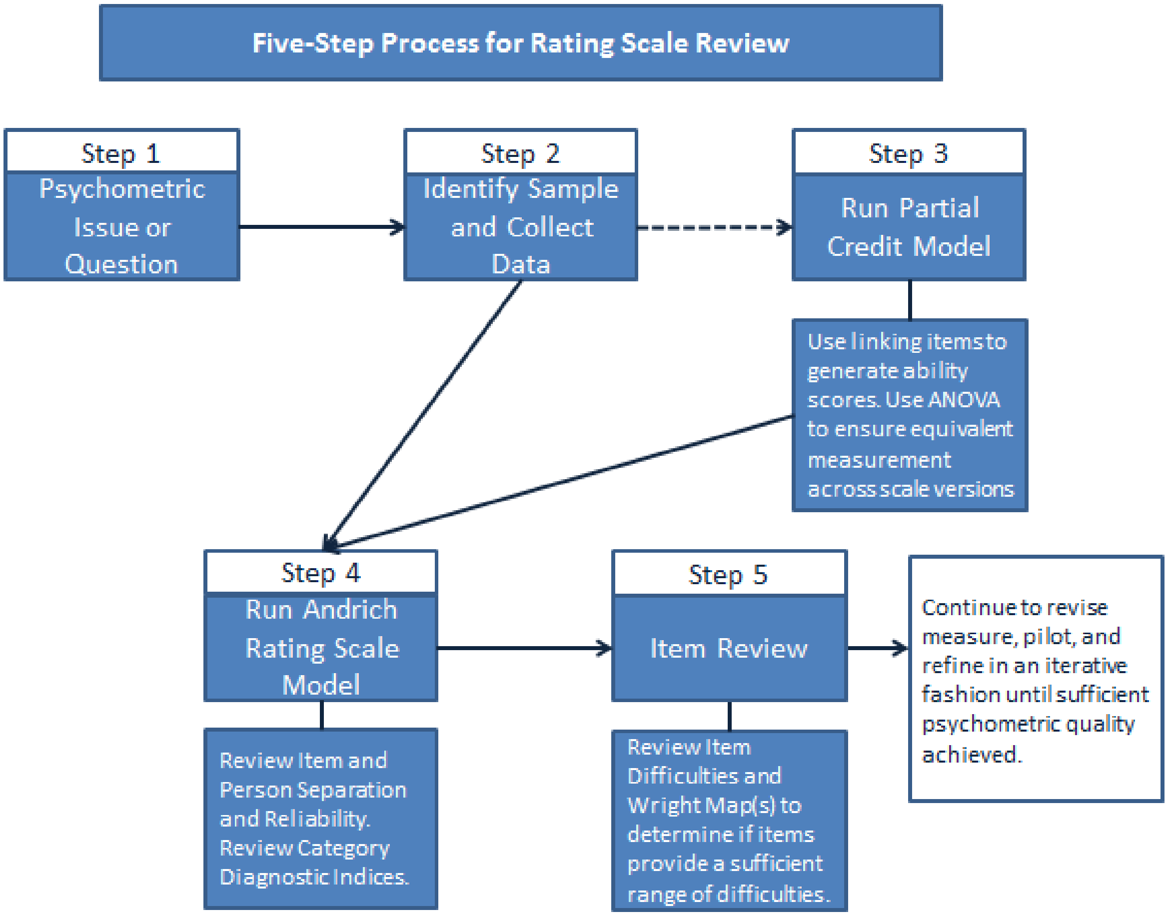
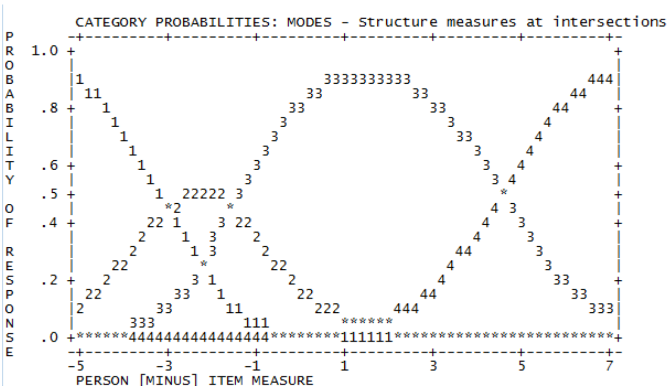
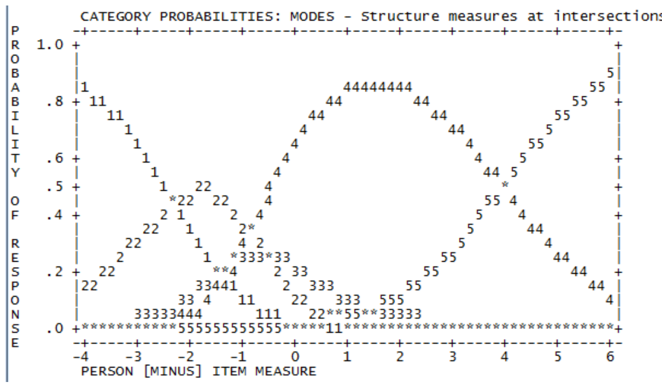


Figure 2: Course Effectiveness -- Probability Curves of Response Categories by Condition

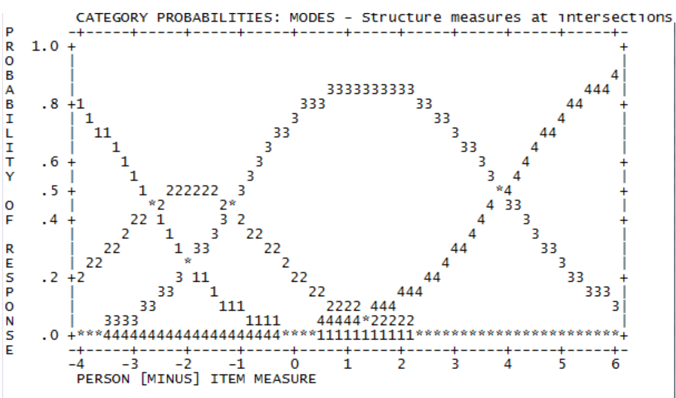
Condition 1



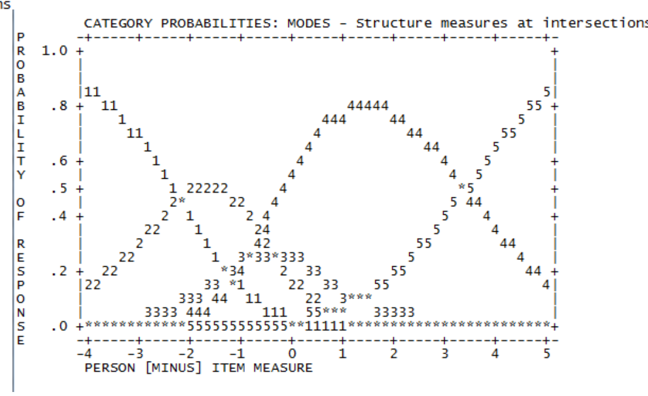
Condition 2



Condition 3[1]



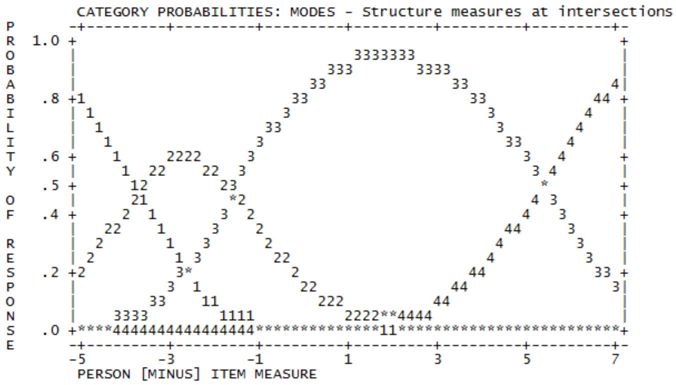
Condition 4[1]



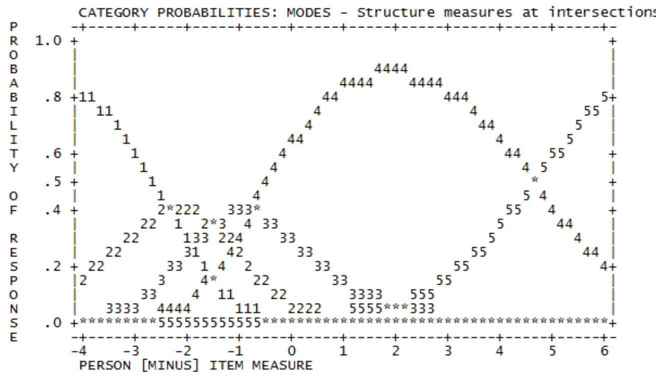
Notes: [1] The RSM was run with the don't know/not applicable option coded as missing data.

Figure 3: Instructional Effectiveness -- Probability Curves of Response Categories by Condition

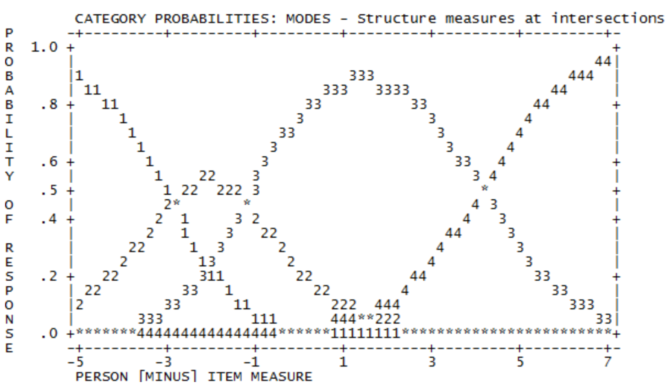
Condition 1



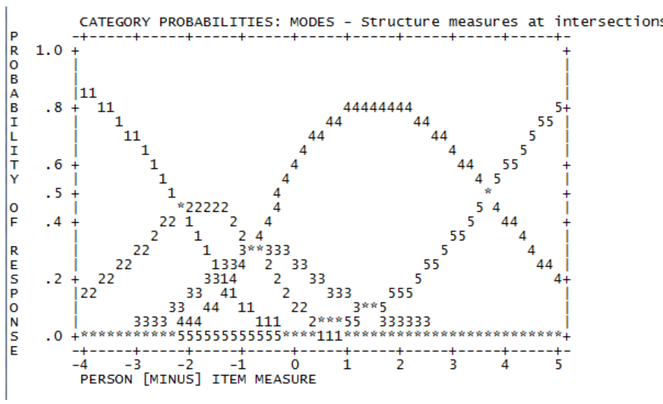
Condition 2



Condition 3[1]



Condition 4[1]



<u>Notes</u>: [1] The RSM was run with the don’t know/not applicable option coded as missing data.

Figure 4: Course Effectiveness – Wright Map Showing Item Difficulty vs. Person Ability

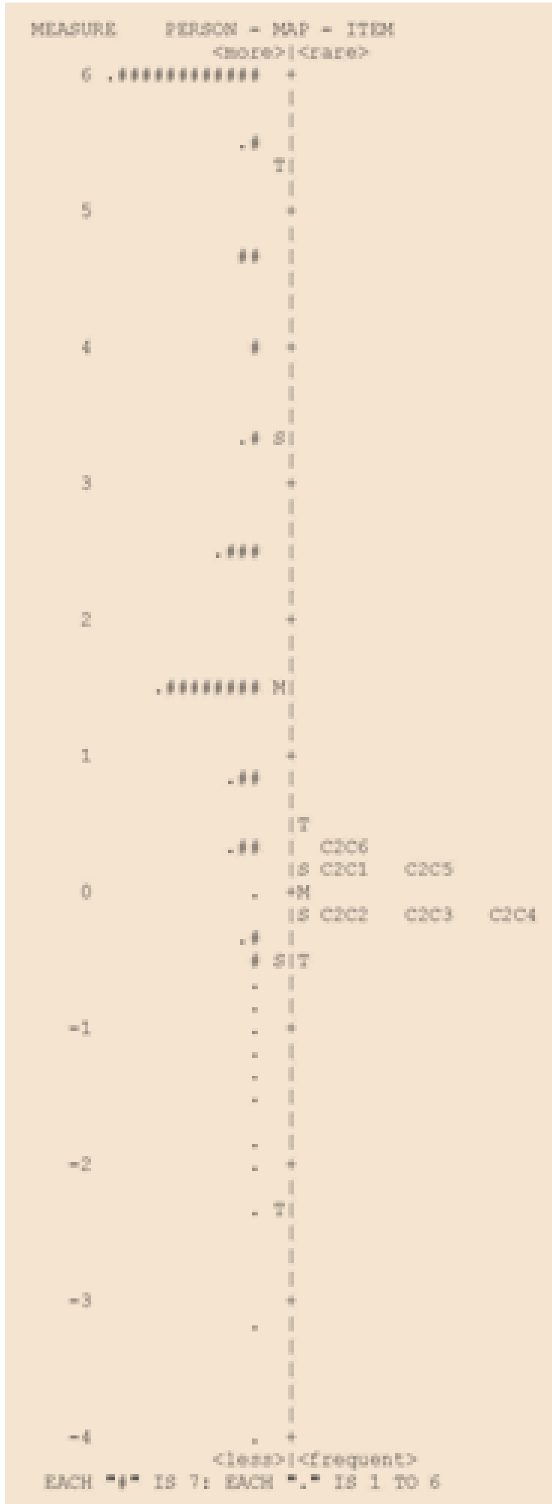


Figure 5: Instructional Effectiveness – Wright Map Showing Item Difficulty vs. Person Ability

Predicting Graduation Outcomes: Identifying Students at Risk of Not Graduating

Meg Munley, Lehigh University

Executive Summary

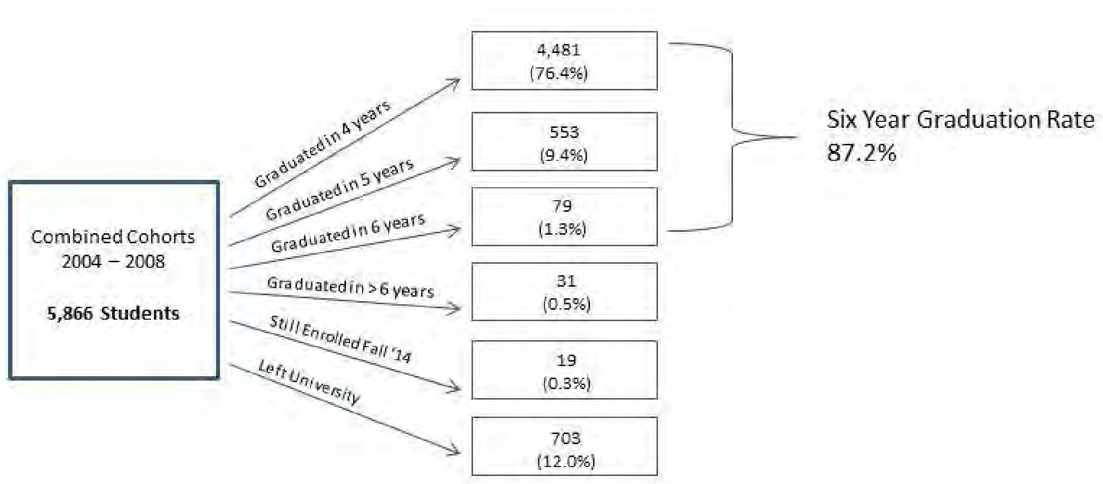
The purpose of this study is to identify factors that affect the likelihood of graduating and to develop a model to predict the likely graduation outcomes of our undergraduate students. By comparing graduating rates across several characteristics, clear differences exist in graduation rates for different groups of students. Looking separately at the likelihood of graduating within six and within four years, a logistic regression analysis was used in order to isolate the effects of certain characteristics on the probability of graduating. The findings show that certain groups of students have an increased likelihood of graduating, controlling for other factors. For instance, women, legacies, varsity athletes, Greek students, and students from the Tri-State area all have an increased likelihood of graduating. An interesting measure of student interest that proved to be a significant predictor of graduating was the admissions contact count. Students with more contacts, which include activities like campus tours and information sessions, have an increased likelihood of graduating. Not surprisingly, academic performance is also a significant predictor of whether or not students graduate. Students with higher first term GPAs and higher rank indexes (a measure of high school grade performance) have an increased likelihood of graduating. Students who have credits which were attempted but not passed during their first term have a decreased likelihood of graduating. Interestingly, the number of credits earned prior to a student's first term proved to be a significant, positive predictor of four year graduation, but not a significant predictor of six year graduation.

This study also demonstrates how the regression model can be used to identify students who may be at risk of not graduating. The regression model uses the student characteristics to estimate a predicted probability of graduating for each student. The accuracy of the model is discussed in detail within the study. Most importantly, there is an element of judgement in how "at risk" is operationally defined. If narrowly defined (i.e., using a lower probability of graduating as the threshold to define "at risk"), a small group of students will be identified. If more broadly defined (i.e., using a higher probability of graduating as the threshold to define "at risk"), a larger group of students will be identified. Using different operational definitions of "at risk" will have consequences on the accuracy of the model. If the model is ultimately used to identify at risk students, it may be useful to compare students identified by this model with other groups of students who have been identified as at risk, such as those on academic probation. Although there may be a large overlap between the lists, it is possible that the list created from the regression model could identify students who may otherwise fall through the cracks.

Introduction

The purpose of this study is to identify factors that affect the likelihood of graduating and to develop a model to predict the likely graduation outcomes of our undergraduate students. This study uses data from the incoming cohorts of 2004 through 2008, the five most recent cohorts for which we can calculate six year graduation rates. The diagram below shows the graduation outcomes of the 5,866 students from these combined cohorts. As shown in the diagram, 76.4% of students graduated within four years and 87.2% graduated within six years. A small percent of students (0.5%) took longer than six years to graduate and a small percent (0.3%) remained enrolled at the institution in the fall of 2014. The remaining 12% of the students left the institution without completing their degree. This study focuses on the likelihood that a student will graduate within the six year time frame.

Figure 1: Graduation Outcomes of the Incoming Cohorts of 2004 through 2008



This study considers several factors that may affect the probability of graduating. These include demographic information (e.g., gender, race/ethnicity), student affiliations (e.g., college affiliation, Greek affiliation), as well as academic performance measures (e.g., SAT scores, first term GPA). Provided below is a full list of student characteristics and academic measures considered in this study.

Characteristics	Academic Measures
Gender	Combined SAT score
Race/Ethnicity	Rank Index (Measure of High School GPA)
Applicant type (Early Decision/Normal Application)	First term GPA
Legacy status	Credit hours attempted but not passed first term
Admissions contact count	Credit hours earned prior to first term
Home state (Tri-State or other)	
Varsity athletic status (during first year)	
Financial measure (institutionally defined gross need)	
College in which student was first enrolled	
Greek affiliation	

Graduation Rates by Student Characteristics

Tables 1 and 2 provide the overall four and six year graduation rates broken down by the student characteristics listed on the previous page. The graduation rates are calculated using all students in the cohorts of 2004 through 2008. In Table 1, graduation rates are broken down by characteristics that are categorical, such as gender and Greek affiliation. In Table 2, graduation rates are broken down by characteristics that are on a continuous scale, like first term GPA and SAT score. For the purpose of the regression analysis described in the following section, the continuous measures are used. For example, a student's actual SAT score is used as a predictor of graduating rather than the SAT range for that score. However, the breakdowns provided in Table 2 may be useful in seeing the general relationship between these continuous measures and graduation rates.

Table 1 shows that female students have higher four and six year graduation rates than male students, with an even greater gap between four year graduation rates. Female and male students had six year graduation rates of 89.9% and 85.2%, respectively. The four year graduation rates for female and male students were 83.6% and 71.2%, respectively. White students also have higher four and six year graduation rates than non-White students. For most race and ethnicity categories, the four year gap is greater than the six year gap. White students had a six year graduation rate of 88.8%. This is 19 percentage points higher than the African American six year graduation rate of 69.8% and 8.5 percentage points higher than the Hispanic graduation rate of 80.3%. The four year graduation rate for White students is 79.0%, which is 28.3 percentage points higher than the African American four year graduation rate of 50.7%. The White four year graduation rate is 13.3 percentage points higher than the Hispanic four year graduation rate of 65.7%.

A very noticeable difference in graduation rates is seen between Greek and non-Greek students. The Greek six year graduation rate is 94.7%, compared to 82.4% for non-Greek students. It is worth noting that students typically join a fraternity or sorority during the spring of their first year. The fact that a student joins a Greek organization may be a strong indication that the student plans to remain at the university (students are unlikely to join if they plan on transferring out of Lehigh). It is also worth noting that there are minimum GPA requirements to join a Greek organization. Students who leave Lehigh for academic reasons, therefore, may not have had the option of joining a fraternity or sorority.

Noticeable differences in graduation rates also exist across other characteristics. For example, legacies, students from the Tri-State area, and varsity athletes have higher graduation rates than their respective counterparts. It is worth noting that in previous analyses, graduation rates have been compared between recruited athletes and students who were not recruited athletes. Although there is an overlap between those who are recruited athletes and varsity athletes, there are students who belong to one group and not the other. While this study includes analysis on varsity athletes instead of recruited athletes, it is interesting to note that for these cohorts of students, the graduation rate for recruited athletes is below the university average and the graduation rate for varsity athletes is above the university average.

Table 1: Graduation Rates by Categorical Student Characteristics

	Count	Percent of Undergraduates	Average 4 Year Graduation Rate	Average 6 Year Graduation Rate
Gender				
Female	2,453	41.8%	83.6%	89.9%
Male	3,413	58.2%	71.2%	85.2%
Race/Ethnicity				
White	4,367	74.4%	79.0%	88.8%
African American	205	3.5%	50.7%	69.8%
Hispanic	274	4.7%	65.7%	80.3%
Asian American	362	6.2%	77.1%	86.5%
Non-resident Alien	177	3.0%	67.2%	83.1%
Two or More Races	137	2.3%	67.9%	81.0%
Other/Unknown	344	5.9%	75.3%	88.4%
Applicant Type				
Early Decision	2,338	39.9%	76.3%	87.6%
Normal Application	3,528	60.1%	76.4%	86.9%
Legacy Status				
Legacy	1,004	17.1%	79.3%	90.9%
Not a Legacy	4,862	82.9%	75.8%	86.4%
Home State				
Tri-State Area (PA, NJ, NY)	3,765	64.2%	78.5%	88.9%
Outside Tri-State Area	2,101	35.8%	72.7%	84.1%
Athletic Status				
Varsity Athlete	913	15.6%	79.1%	89.3%
Not a Varsity Athlete	4,953	84.4%	75.9%	86.8%
Incoming College				
CAS	2,509	42.8%	78.5%	86.3%
CBE	1,270	21.7%	78.1%	87.8%
RCEAS	1,899	32.4%	73.0%	87.8%
Intercollegiate Programs	188	3.2%	70.2%	87.8%
Greek Affiliation				
Greek	2,290	39.0%	83.8%	94.7%
Non-Greek	3,576	61.0%	71.6%	82.4%
Total	**5,866**	**100.0%**	**76.4%**	**87.2%**

Table 2 shows that, not surprisingly, students with higher SAT scores, rank indexes, and first term GPAs are generally more likely to graduate. Students with more gross financial need are generally less likely to graduate. In terms of credit hours, students who have any credits which were attempted but not passed during their first term are less likely to graduate. Students who enter Lehigh with more earned credits, usually through AP credits, are generally more likely to graduate. It is interesting that there is a drop in graduation rate at the very top of the distribution for both SAT scores and credits earned prior to first term. It may also be surprising that there appears to be a strong relationship between the admissions contact count and graduation rates. Contacts include activities such as tours, information sessions, and contacting the admissions office for information (there are many types of contacts). The contact count, therefore, may be a proxy for the student's excitement about Lehigh and the desire to matriculate here. Students with more contacts generally have higher graduation rates.

Table 2: Graduation Rates by Student Characteristics (Continuous Measures)

	Count	Percent of Undergraduates	Average 4 Year Graduation Rate	Average 6 Year Graduation Rate
Admissions Contact Count				
< 5 Contacts	1,371	23.4%	68.1%	82.0%
5 - 9 Contacts	3,567	60.8%	78.1%	88.1%
10+ Contacts	927	15.8%	82.1%	91.2%
Gross Need				
Zero	3,166	54.0%	78.6%	88.6%
$1 - $10,000	281	4.8%	75.4%	85.4%
$10,001 - $20,000	421	7.2%	75.3%	87.6%
$20,001 - $30,000	798	13.6%	78.6%	87.3%
$30,001 - $40,000	897	15.3%	72.1%	84.9%
$40,000 +	303	5.2%	62.4%	79.9%
Combined SAT				
< 1,100	207	3.5%	57.5%	76.3%
1,100 - 1,190	623	10.6%	70.8%	82.5%
1,200 - 1,290	1,555	26.5%	75.3%	86.4%
1,300 - 1,390	2,353	40.1%	79.0%	88.9%
1,400 - 1,490	954	16.3%	79.8%	90.0%
1,500 +	174	3.0%	74.7%	85.1%
Rank Index				
< 66	801	13.7%	64.7%	80.4%
66 - 70	1,925	32.8%	72.7%	85.2%
71 - 75	1,560	26.6%	79.8%	89.0%
76 +	1,572	26.8%	83.9%	91.4%
First Term GPA				
< 2.0	191	3.3%	19.9%	38.7%
2.0 - 2.49	435	7.4%	51.3%	71.7%
2.5 - 2.99	1,307	22.3%	70.3%	86.4%
3.0 - 3.49	1,990	33.9%	82.0%	91.3%
3.5 +	1,943	33.1%	85.9%	91.8%
First Term Credit Hours Attempted but not Earned				
Zero	5,200	88.6%	79.2%	88.9%
Any	666	11.4%	54.8%	73.4%
Credit Hours Earned Prior to First Term				
Zero	2,180	37.2%	68.5%	83.1%
1 - 10	2,096	35.7%	78.9%	88.9%
11 - 20	1,139	19.4%	84.0%	91.1%
21 +	451	7.7%	83.6%	88.9%
Total	**5,866**	**100.0%**	**76.4%**	**87.2%**

Logistic Regression Model

In order to isolate the effects of certain student characteristics on graduation rates, a regression analysis was used on the entire sample of students in the data set. Due to the binary nature of the outcome variable (a student graduates or does not), a logistic regression model was used to estimate the probability of graduating. In this analysis, graduation was modeled as a function of student characteristics listed in the previous section. Logistic regression uses Maximum Likelihood Estimation (MLE) to determine the coefficients that provide the greatest probability of correctly predicting the outcomes in the data set. In a logistic model, the outcome is transformed into the log of the odds ratio. The coefficients represent the unit change in the log of the odds ratio for each unit change in the predictor. In order to make the results *slightly* easier to interpret, the coefficients are transformed by exponentiation. The transformed coefficient, Exp (b), can then be interpreted as how the predictor relates to the odds ratio, instead of the log of the odds ratio.[1] For example, if the exponent of a coefficient is 1.15, then a unit change in this variable increases the odds of graduating by 15%. If the exponent of the coefficient is 0.85, then a unit change in this variable decreases the odds of graduating by 15%.

Table 3: Descriptive Statistics of Predictors

Predictor	Minimum	Maximum	Mean	Std. Dev.
Female	0	1	.42	0.49
African American	0	1	.04	0.18
Hispanic	0	1	.05	0.21
Asian	0	1	.06	0.24
Two or More Races	0	1	.02	0.15
Non-Resident Alien	0	1	.03	0.17
Other/Unknown Race/Ethnicity	0	1	.06	0.23
Early Decision	0	1	.40	0.49
Legacy	0	1	.17	0.38
Admissions Contact Count	1	22	6.69	2.86
From Tri-State Area	0	1	.64	0.48
Varsity Athlete	0	1	.16	0.36
Gross Need	0	$51,259	$12,368	$15,418
First College: CBE	0	1	.22	0.41
First College: RCEAS	0	1	.32	0.47
First College: Interdisciplinary	0	1	.03	0.18
Greek	0	1	.39	0.49
CombinedSAT	890	1,600	1,307	106
Rank Index	47	80	71.36	5.82
First Term GPA	0.00	4.00	3.17	0.59
First Term Credit Hours - Attempted, Not Passed	0	15	.43	1.39
Credit Hours Earned Prior to First Term	0	106	7.00	8.41

[1] Odds of an event: The probability of event occurring divided by the probability of the event not occurring.
Odds ratio (used to compare the odds of two groups): Odds of an event for one group divided by the odds for another group.

Table 3 provides the descriptive statistics of the variables included in the regression analysis.[2] Tables 4 and 5 provide the regression results for six year and four year graduation rates, respectively. Although the focus of this study is predicting whether or not students will graduation within the six year time span, the results indicate that it may be useful to compare the results and follow up with a more in-depth, separate analysis of the time to graduation.

Table 4: Results from Logistic Regression Predicting Graduation within <u>Six Years</u>

Predictors	B	S.E.	Significance (p-value)	Exp(B)
Female	0.32	0.10	0.00	1.38
African American	-0.09	0.21	0.68	0.92
Hispanic	0.00	0.19	0.99	1.00
Asian	0.22	0.17	0.20	1.25
Two or More Races	-0.18	0.26	0.48	0.84
Non-Resident Alien	0.30	0.24	0.21	1.34
Other or Unknown Race/Ethnicity	0.06	0.18	0.72	1.07
Early Decision	-0.03	0.10	0.74	0.97
Legacy	0.26	0.13	0.04	1.30
Admissions Contact Count	0.07	0.02	0.00	1.07
From Tri-State Area	0.42	0.09	0.00	1.52
Varsity Athlete	0.77	0.14	0.00	2.16
Gross Need	0.00	0.00	0.87	1.00
First College: CBE	0.21	0.12	0.08	1.23
First College: RCEAS	0.35	0.11	0.00	1.42
First College: Intercollegiate Program	0.48	0.25	0.05	1.62
Greek	1.30	0.11	0.00	3.68
CombinedSAT	0.00	0.00	0.71	1.00
Rank Index	0.02	0.01	0.01	1.02
First Term GPA	0.92	0.08	0.00	2.52
First Term Credit Hours - Attempted, Not Passed	-0.08	0.03	0.00	0.92
Credit Hours Earned Prior to First Term	0.01	0.01	0.18	1.01
Constant	-3.61	0.85	0.00	0.03

Table 4 provides information on which variables are significant predictors of graduating within six years. The results show that the highlighted characteristics were significant predictors of graduating within six years (at the 5% significance level); predictors that are not highlighted were not significant predictors of graduating within six years. The table also provides the difference in the odds of graduating for students with different characteristics. Again, if the Exp(B) > 1, this means that the predictor has a positive effect on the likelihood of graduating; if Exp(B) < 1, the predictor has a negative effect on the likelihood of graduating. For example, the odds of a female student

[2] The regression analysis was based on 5,856 students (10 students with missing data were excluded from the analysis).

graduating within six years are 38% higher than the odds of a male student graduating within six years, controlling for the other characteristics included in the model. Other characteristics that have a positive effect on the likelihood of graduating include: being a legacy, being a varsity athlete, coming from the Tri-State area, and entering Lehigh within the College of Engineering & Applied Science or one of the intercollegiate programs (compared to entering in the College of Arts & Sciences). Although entering in the College of Business & Economics was not significant at the 5% level, it is significant at the 10% level (p=.08) and the effect on the likelihood of graduating was positive compared to those entering in the College of Arts & Sciences. The results show that while having a higher Rank Index has a positive and significant effect on the likelihood of graduating, the SAT score did not have a significant effect when controlling for other factors. Students who have more credit hours attempted but not passed during the first term have a lower likelihood of graduating within six years. On the other hand, the number of credits earned prior a student's first term was not a significant predictor in the model. The two strongest predictors in the model are first term GPA (Exp(B) = 2.52) and Greek affiliation (Exp(B) = 3.68). This may not be surprising given the noticeable differences in graduation rates displayed in Tables 1 and 2.

Table 5 provides the regression results for the four year graduation model. In terms of which characteristics were significant predictors of graduation, the four and six year models have similar lists of significant predictors, with a few notable differences. In the four year model, there is a negative and significant effect on the likelihood that African American students will graduate compared to White students. This effect is not seen in the six year graduation model. It may be important to recall, from Table 1, that the four year graduation gap between White and African American students is even greater than the six year gap. Another difference is seen in the significance of the number of credit hours earned prior to a student's first term. While not significant in predicting whether or not a student will graduate within six years, the number of credit hours earned prior to the first term is significant in predicting whether or not a student will graduate in four years. The combined SAT score also appears to have a significant effect on the probability of graduating within four years, although the effect size is negligible (Exp(B) = 1.00). There are two variables which are significant in predicting six year graduation but not four year graduation. In the four year model, there is no longer a significant, positive effect of entering into the College of Engineering & Applied Science compared to entering within the College of Arts & Sciences. This may not be surprising considering the lower four year graduation rate for students who enter Lehigh within the College of Engineering & Applied Sciences (see Table 1). Legacy status is another student characteristic that appears to be a significant predictor of six year graduation but is not a significant predictor of four year graduation.

The regression results show that there are some noticeable differences between how student characteristics affect the likelihood of graduating within four years and within six years. While the odds of a woman graduating within six years is 38% higher than the odds of a man graduating within that time frame, the odds for a woman graduating within four years are 69% higher than the odds for a man. This difference in odds is consistent with the larger four year graduation gap between men and women that is shown in Table 1. The results also show a significant difference in the odds of graduating within four years between White and African American students. These differences are worth exploring in a separate, in-depth analysis on the time to graduation. (Stay tuned.)

Table 5: Results from Logistic Regression Predicting Graduation within Four Years

Predictors	B	S.E.	Significance (p-value)	Exp(B)
Female	0.52	0.08	0.00	1.69
African American	-0.39	0.18	0.03	0.67
Hispanic	-0.20	0.16	0.21	0.82
Asian	0.17	0.14	0.25	1.18
Two or More Races	-0.30	0.21	0.16	0.74
Non-Resident Alien	-0.15	0.19	0.44	0.86
Other or Unknown Race/Ethnicity	-0.14	0.14	0.31	0.87
Early Decision	-0.10	0.08	0.19	0.90
Legacy	0.01	0.09	0.91	1.01
Admissions Contact Count	0.07	0.01	0.00	1.07
From Tri-State Area	0.31	0.07	0.00	1.36
Varsity Athlete	0.61	0.11	0.00	1.84
Gross Need	0.00	0.00	0.11	1.00
First College: CBE	0.20	0.09	0.04	1.22
First College: RCEAS	-0.12	0.09	0.15	0.88
First College: Intercollegiate Program	-0.09	0.19	0.61	0.91
Greek	0.64	0.08	0.00	1.90
CombinedSAT	0.00	0.00	0.01	1.00
Rank Index	0.02	0.01	0.00	1.02
First Term GPA	1.04	0.07	0.00	2.83
First Term Credit Hours - Attempted, Not Passed	-0.11	0.02	0.00	0.90
Credit Hours Earned Prior to First Term	0.02	0.01	0.00	1.02
Constant	-3.51	0.69	0.00	0.03

Predicted Percentage Point Change in Probability of Graduating

Although easier to interpret than a change in the *log* of the odds of an event occurring, interpreting a change in the odds is still not very intuitive. As an example, consider two groups with different graduation rates: group A has a graduation rate of 75% and group B has a graduation rate of 50%. The odds of an event occurring are calculated by taking the probability of event occurring and dividing it by the probability of the event not occurring. For group A, the odds are .75/.25 = 3. For group B, the odds are .5/.5 = 1. One might say "the odds of graduating for group A are 3 to 1 and the odds of graduating for group B are 1 to 1." In this case, the odds of graduating are three times greater for group A than for group B. As a percentage, the odds for group A are 300% larger than the odds for group B. While this is a valid way of looking at differences in probability, most people would prefer saying "Group A has a graduation rate that is 25 percentage points higher than Group B's graduation rate".

Because a percentage point change in graduation rates is much more interpretable than a change in odds, an extra step was taken here to calculate the change in predicted graduation rates for students with different characteristics. The equation below was used to calculate the probability of graduating for the average student. The probability could then be calculated for different groups by changing only the one characteristic in question.

$$\text{Probability of Graduating} = \text{Exp}(B_o + B_1X_1 + \ldots) / [1 + \text{Exp}(B_o + B_1X_1 + \ldots)]$$

Table 6 provides the predicted percentage point change in the probability of graduating for students with different characteristics. Note that the percentage point changes were only calculated for characteristics which were found to have significant effects on the likelihood of graduating in four or six years. As an example of the interpretation here, consider the effects of being female. The results indicate that for an average student on other characteristics, being female increases the likelihood of graduating within six years by 2.68 percentage points, compared to being male (say, from an 87.0% graduation rate to an 89.68% graduation rate). Being female would increase the likelihood of graduating in four years by 8.23 percentage points. This is consistent with seeing a greater difference in odds for a female graduating in four years compared to the odds of graduating in six years. It may also be useful to compare the difference in observed graduation rates between women and men with the predicted difference in the probability of graduating. The actual six year graduation gap between women and men is 6.3 percentage points (89.9% six year graduation rate for women; 83.6% six year graduation rate for men). The actual four year graduation gap between women and men is 12.4 percentage points (83.6% four year graduation rate for women; 71.2% four year graduation rate for men). The predicted percentage point differences displayed in Table 6 are calculated by controlling for the other factors in the model. This is why the predicted difference in the probability of graduating will be different than the observed difference in graduation rates.

The results in Table 6 show that the largest differences in the predicted probability of graduating occur in the comparison between students of different first term GPAs. Compared to students earning a 4.0 during their first term, the predicted decrease in the likelihood of graduating in six years is 6.2 percentage points for those earning a 3.0, 18.78 percentage points for those earning a 2.0, and 38.86 percentage points for those earning a 1.0. The effects of different GPAs are even greater on the probability of graduating within four years. Note that, in the regression results in Table 4, it would appear that Greek affiliation had the largest effect on the probability of graduating within six years (largest Exp(B)). These transformed coefficients, however, measure the change in odds for a *single* unit change in the predictor. By comparing multiple unit changes in GPA (from a 2.0 to a 3.0, from a 2.0 to a 4.0), the results in Table 6 show that the first term GPA appears to be the stronger predictor of graduating.

Table 6: Predicted Percentage Point Change in Probability of Graduating

Predictors	Predicted Percentage Point Change in Probability of: Four Year Graduation	Six Year Graduation
Female	+ 8.23	+ 2.68
African American	- 6.99	---
Legacy	---	+ 2.08
Admissions Contact Count		
5 Contacts (Compared to 1 Contact)	+ 5.05	+ 2.73
10 Contacts (Compared to 1 Contact)	+ 10.36	+ 5.42
From Tri-State Area	+ 5.13	+ 3.70
Varsity Athlete	+ 8.63	+ 5.31
First College: CBE	+ 3.05	---
First College: RCEAS	---	+ 2.96
First College: Intercollegiate Program	---	+ 13.62
Greek	+ 9.92	+ 10.18
Rank Index		
70 (Compared to 80)	- 3.71	- 1.65
60 (Compared to 80)	- 8.00	- 3.60
50 (Compared to 80)	- 12.83	- 5.88
First Term GPA		
3.0 (Compared to 4.0)	- 13.60	- 6.20
2.0 (Compared to 4.0)	- 36.55	- 18.78
1.0 (Compared to 4.0)	- 61.23	- 38.86
First Term Credit Hours - Attempted, Not Passed		
3 (Compard to 0)	- 5.43	- 2.13
6 (Compared to 0)	- 11.79	- 4.69
9 (Compared to 0)	- 18.95	- 7.72
Credit Hours Earned Prior to First Term		
5 (Compared to 0)	+ 1.95	---
10 (Compared to 0)	+ 3.77	---
15 (Compared to 0)	+ 5.47	---

Model Fit

While it is useful to know which student characteristics have significant effects on the likelihood of graduating, it is also important to consider the accuracy of the model. In logistic regression, the model determines the coefficients that provide the greatest probability of correctly predicting the outcomes in the data set. These coefficients are used to estimate a probability for each student. Therefore, each student in the data set has a predicted probability of graduating (all probabilities fall between 0 and 1). Because Lehigh has a high graduation rate, most students will have a high predicted probability of graduating. The average predicted probability is 0.872, which matches the overall graduation rate of 87.2%. If the goal is to predict graduation outcomes, a probability threshold needs to be determined in order to classify student by category (predicted to graduate vs. predicted to not graduate). The question here is: below what probability should a student be considered at risk for not graduating? In other words, how should "at risk" be operationally defined? The default threshold in this type of regression analysis is often set to 0.5. This means that those with a predicted probability greater than 0.5 will be predicted to graduate; those with a predicted probability below 0.5 will be predicted to not graduate. Because Lehigh students graduate at a high rate, it is worth exploring the effects of increasing that threshold to, say, 0.6, 0.7, or 0.8.

Figure 2: Visual Representation of Model Fit

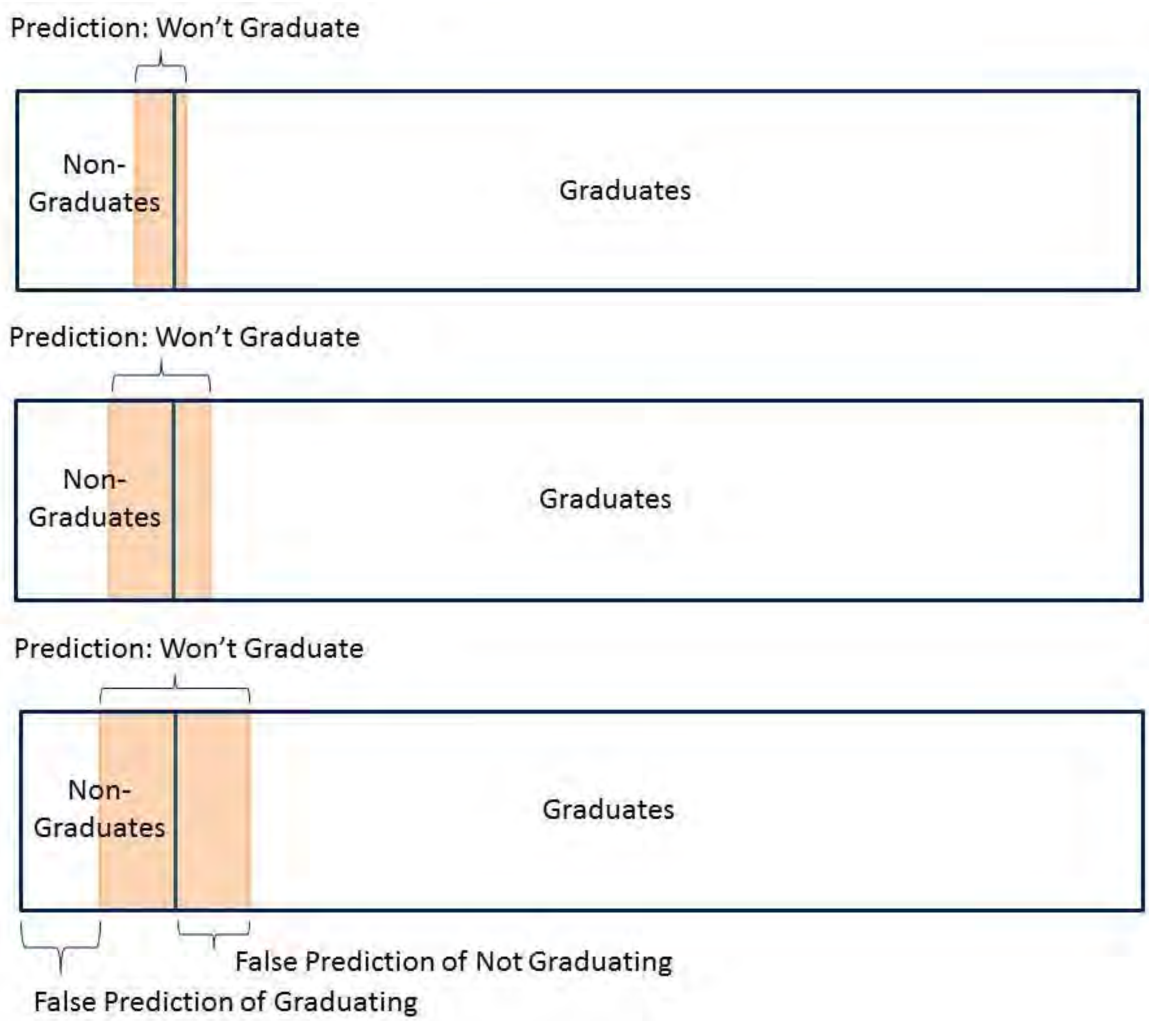


Figure 2 consists of three diagrams that provide a visual representation of the effects of increasing the probability threshold. In the first diagram, a low probability threshold is used (say, 0.5). The model predicts that a relatively small percent of the students will not graduate. This model fails to identify a large portion of non-graduates. In other words, for most of the non-graduates, the model falsely predicts that they will graduate. However, among the students who the model predicts will not graduate (the orange area in the diagram), most are non-graduates. In the second and third diagrams, the probability threshold is increased in order to identify more non-graduates in the predicted non-graduates. This would be analogous to increasing the probability threshold to, say, 0.6 or 0.7. These models do identify more of the non-graduates. In other words, the proportion of non-graduates who have a false prediction of graduating decreases. However, a greater proportion of the predicted non-graduates are actually graduates. This is the trade-off that is faced in the model. If the goal is to identify those who will not graduate, a higher threshold needs to be used to identify a larger percent of non-graduates. That higher threshold, however, means that the percent of predicted non-graduates who are actually non-graduates decreases.

Tables 7 and 8 show how accurately the models predict six and four year graduation outcomes, respectively. Four probability thresholds were used: 0.5, 0.6, 0.7, and 0.8. Two sets of percentages have been highlighted. The percentages highlighted in green answer the question: Among all non-graduates, what percent does the model correctly predict will not graduate? The percentages highlighted in yellow answer the question: Among those who the model predicts will not graduate, what percent actually don't graduate? In terms of these two measures, the four year model is more accurate in predicting the former and the six year model is more accurate in predicted the latter.

For the <u>six year</u> graduation model, the results show:

- Probability threshold of 0.5: Among all non-graduates, the model predicts that 14.4% will not graduate. Among the predicted non-graduates, 71.5% were actually non-graduates.
- Probability threshold of 0.6: Among all non-graduates, the model predicts that 19.0% will not graduate. Among the predicted non-graduates, 62.3% were actually non-graduates.
- Probability threshold of 0.7: Among all non-graduates, the model predicts that 28.7% will not graduate. Among the predicted non-graduates, 49.0% were actually non-graduates.
- Probability threshold of 0.8: Among all non-graduates, the model predicts that 46.9% will not graduate. Among the predicted non-graduates, 34.9% were actually non-graduates.

For the <u>four year</u> graduation model, the results show:

- Probability threshold of 0.5: Among all non-graduates, the model predicts that 23.8% will not graduate. Among the predicted non-graduates, 68.7% were actually non-graduates.
- Probability threshold of 0.6: Among all non-graduates, the model predicts that 35.6% will not graduate. Among the predicted non-graduates, 58.9% were actually non-graduates.
- Probability threshold of 0.7: Among all non-graduates, the model predicts that 50.6% will not graduate. Among the predicted non-graduates, 46.8% were actually non-graduates.
- Probability threshold of 0.8: Among all non-graduates, the model predicts that 70.5% will not graduate. Among the predicted non-graduates, 36.2% were actually non-graduates.

It may be helpful to consider how these measures would translate to a single cohort of students. Because this data set uses five cohorts of students, dividing the student counts in Tables 7 and 8 by five would provide the average outcomes of the regression model. For these cohorts of students, the average cohort consisted of 1171 students. The average number of graduates and non-graduates within six years were 1021 and 150, respectively. Using different graduation probability thresholds, the model would yield the following average results for the six year graduation prediction.

- Threshold of 0.5: Model predicts 30 students will not graduate, 22 of which actually do not graduate.
- Threshold of 0.6: Model predicts 46 students will not graduate, 28 of which actually do not graduate.
- Threshold of 0.7: Model predicts 88 students will not graduate, 43 of which actually do not graduate.
- Threshold of 0.8: Model predicts 201 students will not graduate, 70 of which actually do not graduate.

Table 7: Accuracy of Six Year Graduation Model

Probability Threshold: 0.5 (Those with a probability of below 0.5 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	5064	43	5107	99.2%
Did Not Graduate in 6 Years	641	108	749	14.4%
Total	5705	151	5856	
Percent Correct	88.8%	71.5%		88.3%
Probability Threshold: 0.6 (Those with a probability of below 0.6 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	5021	86	5107	98.3%
Did Not Graduate in 6 Years	607	142	749	19.0%
Total	5628	228	5856	
Percent Correct	89.2%	62.3%		88.2%
Probability Threshold: 0.7 (Those with a probability of below 0.7 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	4883	224	5107	95.6%
Did Not Graduate in 6 Years	534	215	749	28.7%
Total	5417	439	5856	
Percent Correct	90.1%	49.0%		87.1%
Probability Threshold: 0.8 (Those with a probability of below 0.8 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	4452	655	5107	87.2%
Did Not Graduate in 6 Years	398	351	749	46.9%
Total	4850	1006	5856	
Percent Correct	91.8%	34.9%		82.0%

Dividing the student counts in Table 8 by five would provide the average effects of the four year regression model. Again, the average cohort consisted of 1171 students. The average number of graduates and non-graduates within four years were 895 and 276, respectively. Using different graduation probability thresholds, the model would yield the following average results for the four year graduation prediction.

- Threshold of 0.5: Model predicts 96 students will not graduate, 66 of which actually do not graduate.
- Threshold of 0.6: Model predicts 167 students will not graduate, 98 of which actually do not graduate.
- Threshold of 0.7: Model predicts 299 students will not graduate, 140 of which actually do not graduate.
- Threshold of 0.8: Model predicts 538 students will not graduate, 195 of which actually do not graduate.

Note that although the results are presented for a probability threshold of 0.8, this threshold would not make sense for practical purposes because the four year graduation rate is below 80%. For the four year graduation model, over half of the students would be predicted to not graduate at this probability threshold.

Table 8: Accuracy of Four Year Graduation Model

Probability Threshold: 0.5 (Those with a probability of below 0.5 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	4325	150	4475	96.6%
Did Not Graduate in 6 Years	1052	329	1381	23.8%
Total	5377	479	5856	
Percent Correct	80.4%	68.7%		79.5%
Probability Threshold: 0.6 (Those with a probability of below 0.6 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	4132	343	4475	92.3%
Did Not Graduate in 6 Years	890	491	1381	35.6%
Total	5022	834	5856	
Percent Correct	82.3%	58.9%		78.9%
Probability Threshold: 0.7 (Those with a probability of below 0.7 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	3679	796	4475	82.2%
Did Not Graduate in 6 Years	682	699	1381	50.6%
Total	4361	1495	5856	
Percent Correct	84.4%	46.8%		74.8%
Probability Threshold: 0.8 (Those with a probability of below 0.8 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	2759	1716	4475	61.7%
Did Not Graduate in 6 Years	408	973	1381	70.5%
Total	3167	2689	5856	
Percent Correct	87.1%	36.2%		63.7%

Model Validation: 2009 Incoming Cohort

The purpose of creating this model is to identify those who are at risk of not graduating. This requires applying the regression results to a separate set of students in order to estimate predicted probabilities of graduating. Before applying this model to current or future students, it is important to test the model on another set of students. The purpose of testing the model would be to determine whether or not the model is as accurate (or almost as accurate) as it is with the data upon which the model is based. Here, the six year graduation model is used to predict the likely graduation outcomes of the incoming cohort of 2009. Although the six year graduation outcomes for this cohort will not be determined until May 2015, the five year outcomes can be used as a close substitute for the six year graduation outcomes (not many students take a sixth year to graduate).

The incoming cohort of 2009 consisted of 1193 students, 1009 of which graduated within five years. There were 184 students who did not graduate within five years. Table 9 shows how accurate the model was when applying the six year regression model to this cohort of students. Using different graduation probability thresholds, the model yielded the following results.

- Threshold of 0.5: Model predicted 31 students would not graduate, 24 of which actually did not graduate.
- Threshold of 0.6: Model predicted 54 students would not graduate, 40 of which actually did not graduate.
- Threshold of 0.7: Model predicted 108 students would not graduate, 65 of which actually did not graduate.
- Threshold of 0.8: Model predicted 228 students would not graduate, 102 of which actually did not graduate.

By comparing results at different probability thresholds, the previously discussed trade-off in accuracy is apparent. At the low probability threshold of 0.5, the model did not identify many students as predicted non-graduates. The model only predicted that 31 students would not graduate while 184 students actually did not graduate. However, among the 31 students who the model predicted would not graduate, 24 students (77%) actually did not graduate. When the probability threshold is increased, more non-graduates are included in the predicted non-graduates. However, there is also a greater proportion of graduates included in the predicted non-graduate group. For example, using the threshold of 0.7, the model predicted that 108 students would not graduate. Among that group, 65 students (60%) actually did not graduate. Overall, the results show that the model behaved quite similarly on the Cohort of 2009 as it had on the combined cohorts of 2004 through 2008. Moving forward, this is encouraging because it means that the model may have similar accuracy if applied to current or future students.

Table 9: Accuracy of the Six Year Graduation Model Applied to Five Year Graduation Outcomes (Cohort of 2009)

Probability Threshold: 0.5 (Those with a probability of below 0.5 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	1002	7	1009	99.3%
Did Not Graduate in 6 Years	160	24	184	13.0%
Total	1162	31	1193	
Percent Correct	86.2%	77.4%		79.5%
Probability Threshold: 0.6 (Those with a probability of below 0.6 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	995	14	1009	98.6%
Did Not Graduate in 6 Years	144	40	184	21.7%
Total	1139	54	1193	
Percent Correct	87.4%	74.1%		78.9%
Probability Threshold: 0.7 (Those with a probability of below 0.7 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	966	43	1009	95.7%
Did Not Graduate in 6 Years	119	65	184	35.3%
Total	1085	108	1193	
Percent Correct	89.0%	60.2%		74.8%
Probability Threshold: 0.8 (Those with a probability of below 0.8 are predicted not to graduate)				
	Prediction			
	Graduate	Not Graduate	Total	Percent Correct
Observed				
Graduated in 6 Years	883	126	1009	87.5%
Did Not Graduate in 6 Years	82	102	184	55.4%
Total	965	228	1193	
Percent Correct	91.5%	44.7%		63.7%

Discussion and Potential Impact

This study used a logistic regression analysis to estimate the effects that certain student characteristics have on the likelihood that a student will graduate. The regression model can be used to predict the graduation outcomes of current or future students. For future students, it may be preferable to update the data set with the most current data. For instance, the six year graduation outcomes for the incoming cohort of 2009 will be available in May 2015. The model could be updated by including those students in the data set (and perhaps excluding the oldest cohort of 2004).

The model estimates a graduation probability for every student. By reviewing the accuracy of the presented models, it is evident that if a small group of students with the lowest probabilities of graduating is identified as "at risk", most of the students in that small group will truly be at risk of not graduating. If a larger group of students that includes those with slightly higher graduation probabilities is targeted as "at risk", more students who are truly at risk will be targeted. In this larger group, however, more students who would ultimately graduate would also be targeted. This is important to consider when thinking about the potential impact of using this model to identify at risk students.

While this study stops short of suggesting specific interventions for these students, the potential impact of an intervention can be estimated. For this purpose, the incoming cohort of 2009 is used as an example. In this cohort, there were 1193 students. If the smallest group of students was targeted as "at risk", using the probability threshold of 0.5, 31 students would be targeted for an intervention. In this case, resources would be spent on 7 students who would have graduated anyway. The other 24 students could potentially benefit from the intervention and graduate. If all 24 graduate, the overall six year graduation rate for this cohort would be increased by 2 percentage points (24/1193). Under a more reasonable assumption that half of the students might benefit and ultimately graduate, the six year graduation rate for this cohort would increase by 1 percentage point (12/1193). If more students were targeted as "at risk", the potential impact on graduation rate increases. For example, if anyone with a graduation probability of below 0.7 was targeted, 108 students would be included in the "at risk" group. In this case, resources would be spent on 43 students who would have graduated anyway (although they may still benefit in other ways from an intervention). If the other 65 students benefited from the intervention and ultimately graduated, the six year graduation rate for this cohort would increase by 5.4 percentage points (65/1193). Again, under a more reasonable assumption that only half of those students would graduate, the six year graduation rate for this cohort would increase by 2.7 percentage points (32/1193). As an alternative to selecting "at risk" students by a certain probability threshold (below 0.5, 0.6, etc.), students could also be targeted by simply identifying those with the lowest probabilities. For instance, if it was determined that there were enough resources to spend on 75 students, the 75 students with the lowest probabilities could be identified (the probability threshold would fall somewhere between 0.6 and 0.7 in this case).

An important consideration here is that it is unknown which non-graduates would be the most likely to benefit from an intervention. Because this model is heavily driven by first term academic performance, those with the lowest graduation probabilities are often those struggling the most academically. Perhaps these students are the least likely to graduate even with an intervention. On the other hand, perhaps these students are the most likely to benefit from an intervention and ultimately graduate.

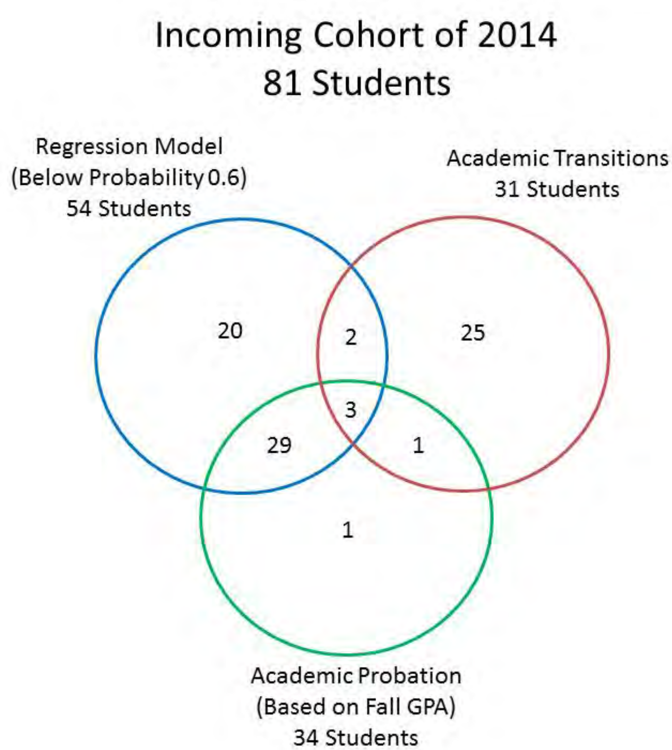
Given that the model is heavily driven by first term academic performance, one may ask whether or not this model is any better at predicting non-graduates than simply identifying those with the lowest first term GPAs. Although the model presented here does a better job at identifying non-graduates, it is not exceptionally better.

Using the 2009 cohort, the accuracy of identifying non-graduates was compared between the model presented here and one using only first term GPAs. Among the 100 students with the lowest estimated probabilities of graduating, 63 did not graduate within five years. Among the 100 students with the lowest first term GPAs, 56 did not graduate within five years. For practical purposes, it may be useful to compare the list created by the regression model with other lists of targeted students, such as those on academic probation. Although there may be a large overlap between the lists, it is possible that the list created from the regression model could identify students who may otherwise fall through the cracks. Again, while this study stops short of making recommendations about what interventions may help at risk students, it is the hope that the findings presented here result in a better understanding of what affects the likelihood of graduating and that the model might be used to identify students at risk of not graduating.

Predicting Graduation Outcomes: Identifying Students at Risk of Not Graduating

Addendum 1: Comparing lists of "at risk" students

As stated in the discussion section of the original report, it may be useful to compare the list of "at risk" students created by the regression model with other lists of targeted students. It is possible that the list created from the regression model could identify students who may otherwise fall through the cracks. Based on conversations with certain Lehigh administrators and staff members, a probability threshold of 0.6 was used to operationally define "at risk" (students with a predicted probability of graduating below 0.6 were considered "at risk"). Using this definition, the regression model identified 54 students as "at risk" in the incoming cohort of 2014. This list was compared with two separate lists of students: those on academic probation in the spring of 2015 and those identified by the Admissions Office as potentially needing extra assistance in their transition to Lehigh. Below, the diagram shows the overlap between the three lists. Between the three lists of students, a total of 81 students have been identified. The regression model identifies 20 students who have not been identified in the other two lists.



Predicting Graduation Outcomes:
Identifying Students at Risk of Not Graduating

Addendum 2:
Can we identify "at risk" students before they arrive on campus?

Several staff members have inquired whether or not this model can be used to identify "at risk" students before they arrive on campus. To do this, the following measures from the regression model would have to be excluded: first term GPA, credits attempted but not passed during the first term, and Greek affiliation. These measures would be excluded because the information does not become available until the spring semester of a student's first year. The regression results are provided below.

By comparing the accuracy of the original model with the accuracy of the pre-matriculation model, it is clear that the model is *much* less accurate when excluding the first term variables. This is not surprising since the two strongest predictors in the original model were first term GPA and Greek affiliation. The model that excludes the first term measures predicts that almost everyone will graduate. If the same criteria are used to identify students who are "at risk" (predicted probability of graduating < 0.6), the original model predicts that 228 students would not graduate. Given that the data span five years, this averages to 46 students per cohort (28 out of those 46 actually do not graduate). In the pre-matriculation model, only 38 students are predicted to not graduate. Given that the data span five years, this averages to just fewer than 8 students per cohort (2 to 3 students out of those 8 actually do not graduate). Again, this is <u>significantly less accurate</u> than the original model. <u>It is not recommended that this model be used to identify students before they matriculate</u>.

Predictors	B	S.E.	Significance (p-value)	Exp(B)
Female	0.46	0.92	0.00	1.58
African American	-0.70	0.19	0.00	0.50
Hispanic	-0.45	0.17	0.01	0.64
Asian	-0.11	0.17	0.51	0.90
Two or More Races	-0.37	0.23	0.11	0.69
Non-Resident Alien	-0.03	0.22	0.90	0.97
Other or Unknown Race/Ethnicity	0.06	0.18	0.72	1.07
Early Decision	-0.07	0.09	0.45	0.93
Legacy	0.34	0.12	0.01	1.40
Admissions Contact Count	0.07	0.02	0.00	1.08
From Tri-State Area	0.34	0.08	0.00	1.41
Recruited Athlete	0.15	0.12	0.23	1.16
Gross Need	0.00	0.00	0.11	1.00
First College: CBE	0.24	0.11	0.03	1.27
First College: RCEAS	0.21	0.10	0.04	1.23
First College: Intercollegiate Program	0.24	0.24	0.31	1.28
CombinedSAT	0.00	0.00	0.65	1.00
Rank Index	0.04	0.01	0.00	1.04
Credit Hours Earned Prior to First Term	0.01	0.01	0.02	1.01
Constant	-2.25	0.12	0.01	0.11

IR Practice: Using Analysis to Drive Decisions in Improving Retention

Shuang Liu
Senior Director of Institutional Effectiveness
Goucher College

Table of Contents

An Executive Summary of Retention Data Analysis

Introduction

The subject of college student retention has captured much attention during the last four decades. Research in this area highlights the complex and multi-faceted relationship between student pre-college characteristics, student expectations, external support, and student academic and social integration at college in relation to retention. The relevant literature generally indicates that the largest variable in predicting student retention is student engagement. As illustrated in *How College Affects Students* (Pascarella and Terenzini, 2005), after reviewing approximately 2,500 studies on college students from the 1990s, and more than 2,600 studies from 1970 to 1990, the authors concluded that student engagement is a central component of student learning and success. In addition, retention and graduation rates are the leading measures of institutional effectiveness and accountability.

Method

To identify key variables in predicting first-year retention rates at Goucher College, as part of the effort of the Retention Data and Analytics Group under the leadership of the Senior Vice President for Strategic Initiatives, the Office of Institutional Effectiveness conducted a multivariate analysis using the most recent five cohorts' record level data (2009-2013). In the dataset, a total of 1,969 records were included. The data suggests that over the past five years, 1,610 out of 1,969 first-time students returned for their Sophomore year, which yielded an average first-year retention rate of 82 percent. Conversely, a total of 359 students transferred out or withdrew from Goucher during their first-year, yielding a five-year average attrition rate of 18 percent.

The dependent variable in the statistical analysis is first-year retention (retained =1, not retained = 0). There are 22 independent variables including demographic characteristics such as gender, ethnicity, age, legacy status, estimated family contribution, and geographical location; student incoming academic abilities such as SAT math and verbal scores, high school Grade Point Average (GPA), math placement, and writing placement; college engagement variables such as student participation in the early immersion program, resident, student athletes, number of credits taken in the fall; and college academic standing such as Fall GPA, spring GPA, and first-year GPA.

Due to the dichotomous nature of the dependent variable, logistic regressions were estimated to account for the predictive relationship between the independent and dependent variables. The independent variables were used as predictors to predict the dichotomous outcome: student returned or not. A predictive model was built where six variables were identified as statistically significant predictors. The classification table of the statistical program used for this analysis suggests that if we use this model to predict student retention, we would be correct 82 percent of the time. In addition, in a regression analysis, multicollinearity arises when there is an extremely high correlation between two or more independent variables in the model. Therefore, the composite SAT score was entered in the equation instead of math and verbal score variables separately. Spring GPA and First Year GPA were not entered in the equation.

Results

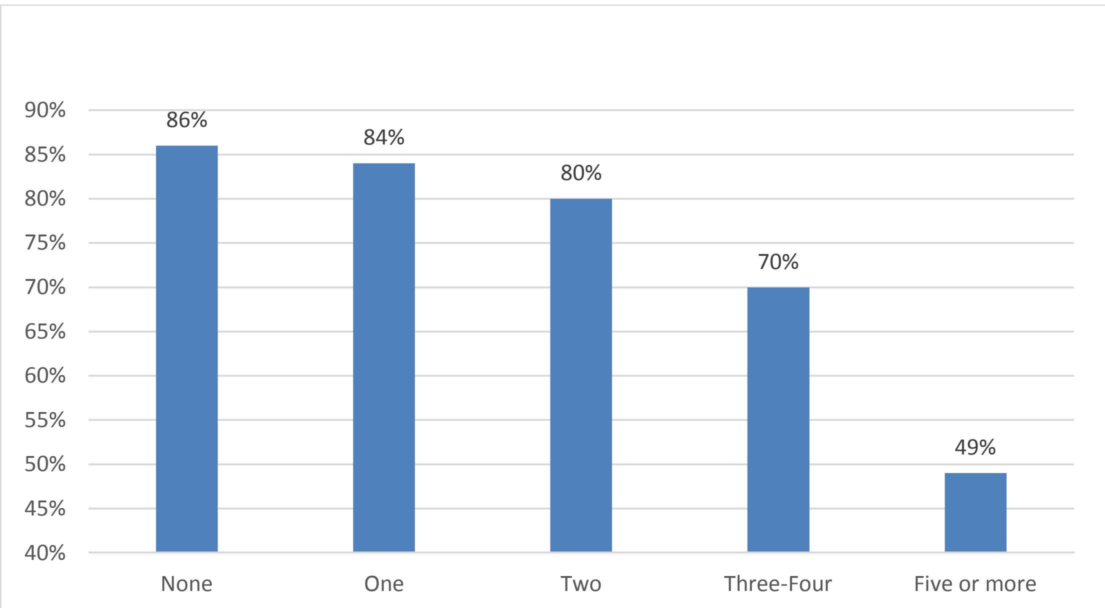
The results of the analysis (Table 1) indicate that the largest variable in predicting student retention at Goucher College is the Fall GPA ($p < 0.01$, meaning that this is a strong predictor). The odds ratio [Exp(B)] suggests that for every one point increase in the fall GPA variable, the odds of a student returning to the College for their Sophomore year increases 1.5 times. Other statistically significant, positive predictors associated with retention include student participation in the early immersion program, participation in student athletics, and the number of credits taken by students in the fall. On the other hand, student age and numbers of reports of concern in the fall were found to be negatively associated with retention. In addition, students who had participated in the Educational Opportunity Program (EOP) were found to be four times likely than the non-EOP students to return after the first-year, despite the disadvantageous variables EOP students tend to be associated with upon college entry. The EOP variable did not appear as a statistically significant variable in the regression model due to the small number of students in the program. Further, high school GPA and math placement variables became statistically significant after college variables were removed.

Table 1. Fall 2009-13 Cohorts Logistic Regression Output

	B	S.E.	Sig.	Exp(B)
Fall GPA	.421	.134	.002	1.523
Age	-.471	.167	.005	.624
Early Immersion	.544	.266	.041	1.724
Student Athletes	.418	.208	.044	1.519
Total Fall Credits	.162	.082	.047	1.176
Number of Reports of Concern in the Fall	-.191	.129	.050	.826

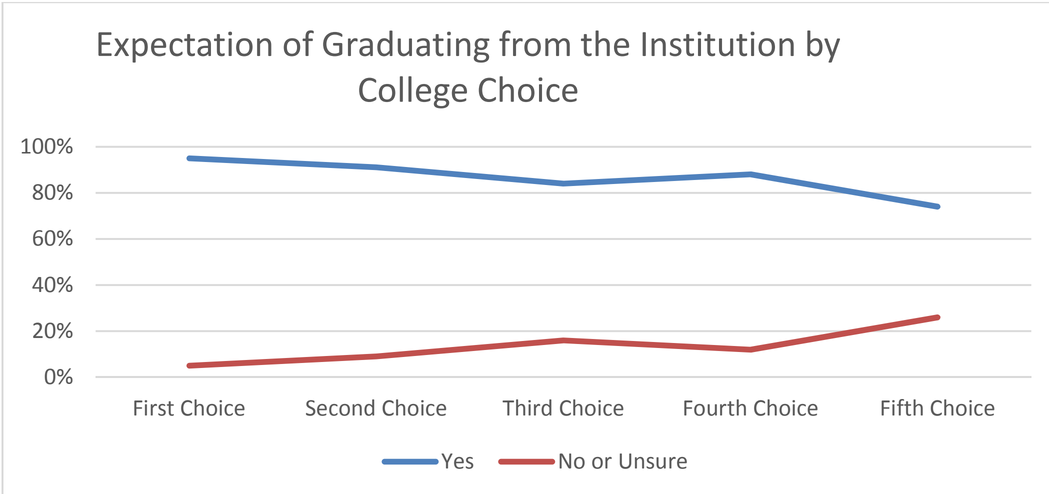
As indicated in Table 1, the number of reports of concern for a student during the fall semester was identified as a statistically significant predictor for retention. A strong correlation exists between numbers of concerns reported in the fall and student retention behavior. The odds ratio [Exp(B)] suggests that for every one increase in the number of reports of concern variable, the odds are 0.83 times as likely for a first-year student to return to the next year. Chart 1 illustrates different retention rates by group based on the value of the reports of concern variable. Receiving three such reports for a student in the fall indicates the critical point of severe alert. In the five-year dataset, 18 percent of the Goucher's first-year students received three or more reports of concern in the fall semester.

Chart 1. Fall 2009-13 Cohorts Retention Rate by Number of Reports of Concern for a Student in the Fall



In addition to the retention data analysis for the most recent five cohorts, the Office of Institutional Effectiveness also had the opportunity to examine the relationship between college expectations captured in the Beginning College Survey of Student Engagement (BCSSE) and student retention. The BCSSE contains a variety of questions related to student pre-college experiences and college expectations and attitudes toward the first-year experiences. Due to the recent revision of the BCSSE instrument, combined multi-year data was not available for access. Based on the Fall 2013 BCSSE data, a correlation was found between a student's choice of the institution and the student's expectation to graduate from this institution (Chart 2).

Chart 2. Fall 2013 Beginning College Survey of Student Engagement Survey Item



Students who initially displayed a lack of institutional commitment were found to be more likely to withdraw. Approximately 40 percent of the students who initially expressed no plan to graduate from Goucher (13 out of 34 students in the fall 2013 cohort) did not return for the Sophomore year.

Since the Fall GPA variable was identified as the most significant variable in predicting student retention in the aforementioned analysis, a multiple regression analysis was conducted to examine the relationship between BCSSE variables with Fall GPA. Learning strategies and the importance of campus environment scores captured in BCSSE were identified as statistically significant variables that contribute to Fall GPA. Construction of a logistic regression model was attempted; however, none of the BCSSE variables were found to directly contribute to the first-year retention rate.

Table 2. Spring 2014 National Survey of Student Engagement (NSSE) First-year Engagement Scores by Retained and Not Retained Students

NSSE Engagement Indicators		N	Mean	Std. Deviation	Std. Error Mean	Median
High-order Learning	Not Retained	9	36.11	12.94	4.31	35.00
	Retained	96	41.72	12.73	1.30	40.00
Reflective and Integrative Learning	Not Retained	9	30.16	11.17	3.72	34.29
	Retained	102	40.99	11.77	1.17	40.00
Learning Strategies	Not Retained	8	33.33	7.13	2.52	33.33
	Retained	92	41.88	13.42	1.40	40.00
Quantitative Reasoning	Not Retained	9	25.93	22.22	7.41	20.00
	Retained	100	25.47	14.54	1.45	23.33
Collaborative Learning	Not Retained	11	26.82	11.89	3.58	25.00
	Retained	106	36.42	12.42	1.21	35.00
Discussion with Diverse Others	Not Retained	8	51.88	8.84	3.13	52.50
	Retained	95	42.79	13.58	1.39	40.00
Student-Faculty Interaction	Not Retained	9	21.11	12.19	4.06	20.00
	Retained	98	22.60	14.43	1.46	20.00
Effective Teaching Practices	Not Retained	10	38.60	13.79	4.36	36.00
	Retained	99	43.54	10.01	1.01	44.00
Quality of Interaction	Not Retained	8	38.31	7.71	2.73	36.75
	Retained	96	43.96	9.05	0.92	44.50
Support Environment	Not Retained	8	37.19	15.44	5.46	36.25
	Retained	94	38.43	11.53	1.19	38.75

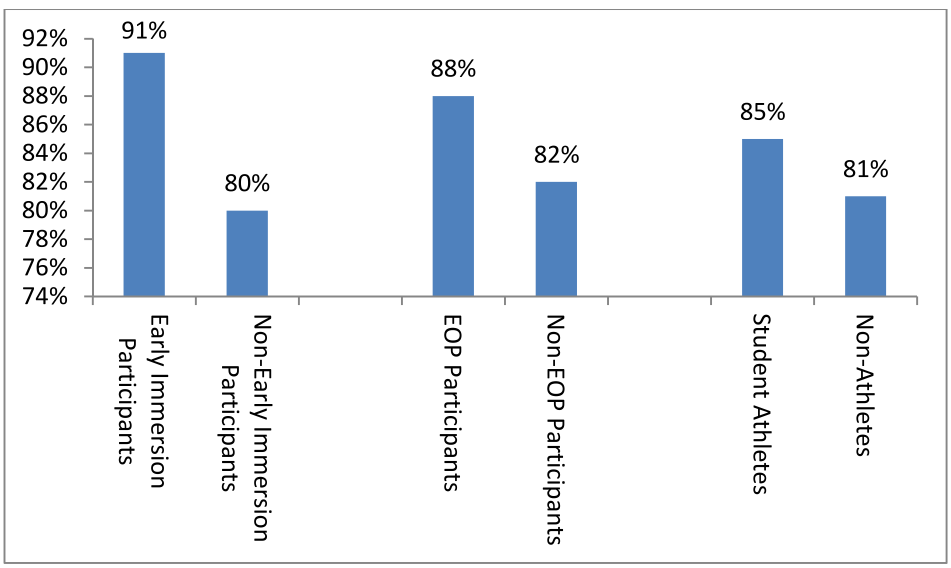
Further, the literature consistently points out that student engagement is the single largest variable in predicting student retention. To identify the different engagement patterns between the students who returned and who did not, the raw data from the 2014 the National Survey of Student Engagement (NSSE) was merged with the retention data. The NSSE measures the amount of time and effort students put into their studies and other educationally purposeful activities. Table 2 lists the comparative results of the 10 NSSE engagement indicators between the students who persisted and who did not. Since NSSE was administered in the Spring semester, the students who transferred out or withdrew from the College by the end of the fall semester did not have the opportunity to participate in the survey. Although the small sample in the data file limits statistical procedures, the Office of Institutional Effectiveness found that returning students in general reported a higher level of engagement, particularly on the indicators of collaborative learning, learning strategies, effective teaching practices, and quality of interaction.

Implications

The analytical results provide rich implications in both retention practice on campus and in future retention data analysis. Specifically, college academic performance seems to be the most important variable contributing to first year student success measured by student retention. Academic support services and the one-on-one connection between students and their advisors/mentors are the essential ingredient to this success. To dig deeper on student academic performance, the Office of Institutional Effectiveness further identified the courses what students are mostly likely to fail in their first-year. The results could help inform placement and tutoring and other academic support services on campus.

Given the fact that the number of reports of concern for a student during the fall semester is a statistically significant predictor for retention, faculty and academic advisors play an important role in identifying at-risk students. Student support staff, including academic support and student development areas, play an important role in following up with these students to ensure intervention programs are effectively delivered. Student access to support services needs to be recorded and analyzed, not only for obtaining the longitudinal data record for the student, but also for continuous improvement and assessment at a program level.

Chart 3. Fall 2009-13 Cohorts First-year Retention Rate by Group



Based on the finding that student athletes, EOP students, and students who participated in the early immersion program are associated with a higher retention rate than their peers (Chart 3), further insight could reveal how to extend a high level of engagement to the entire first-year student population and whether some or all of the strategies employed with these groups fit into a concept to generate tangible results on a large scale.

As for future data collection to better inform retention practice on campus, it is recommended that we should a) fully take advantage of the BCSSE and NSSE data, and b) adopt a comprehensive data system capturing student academic and social integration on campus. The BCSSE Student Advising Report, an individualized student level report, can help faculty and staff gain more knowledge on a student, identify potential issues and connect the student with different types of programs and activities on campus. The BCSSE Student Advising Report should be added to the academic advising tool box. In addition, intervention programs can be designed and delivered to address group issues brought to the surface by the BCSSE. For instance, workshops or seminars can be conducted to the students who need assistance in developing learning strategies at college or additional attention might need to be given to the students who seemed to have a lower level of institutional commitment upon college entry.

Given the fact that student engagement positively contributes to student retention and success, the NSSE survey needs to be locally analyzed and results should be shared on campus. Important variables such as major, student athletic participation, participation in the Frontiers program, residence hall, and other programs that students are affiliated with or engaged in should be coded in the survey. In this way, student engagement indicators can be analyzed and reported at a program level. Significant findings should be shared with the academic and relevant administrative departments to sustain and enhance our strengths, as well as to make improvements in those aspects that may potentially challenge us.

Lastly, in order to ensure that future retention intervention strategies are developed based on diagnostic and constructive data analysis and interpretation, more comprehensive data elements of student engagement need to be collected, including student participation in campus organizations, student clubs, and other deliberate retention intervention programs. The purpose of conducting data analysis on previous student cohorts is to not only report, but more importantly to forecast and intervene with retention behaviors of current students. More comprehensive data will better enable us to do so.

The ongoing efforts in the Retention Data and Analytics Group promote an institutional culture where data is used to inform decision-making and policy development. The collective commitment among the administrators, faculty, and students in the group is evident and encouraging. The Office of Institutional Effectiveness will continue to provide quality information and analytical services to support the College's strategic initiative of improving student retention. Any questions pertaining to this summary can be addressed to Shuang at Shuang.liu@goucher.edu .

Appendix

Faculty Retreat – Round Table Discussion
Topic: First-year Student Retention
Facilitator: Shuang Liu

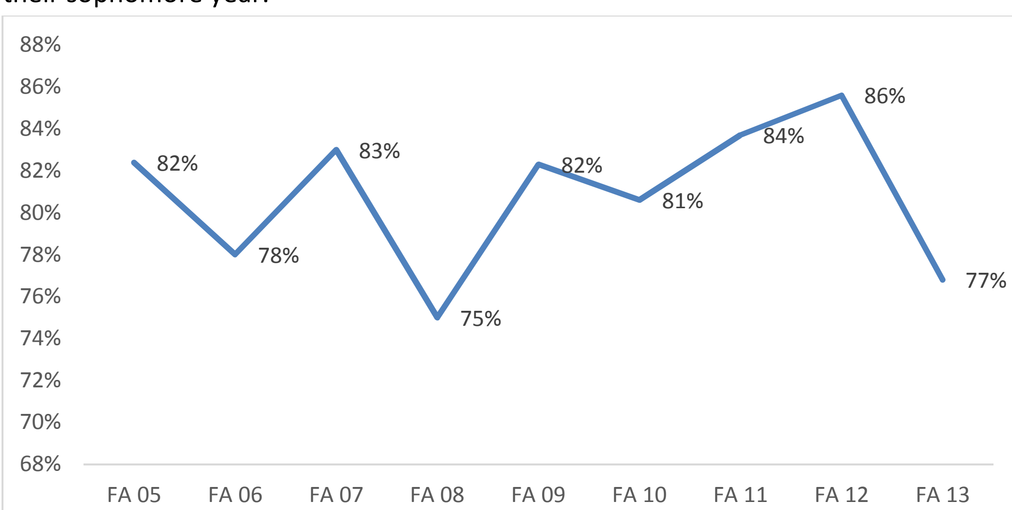
Facts and Figures about Goucher's Retention Data

Background Information

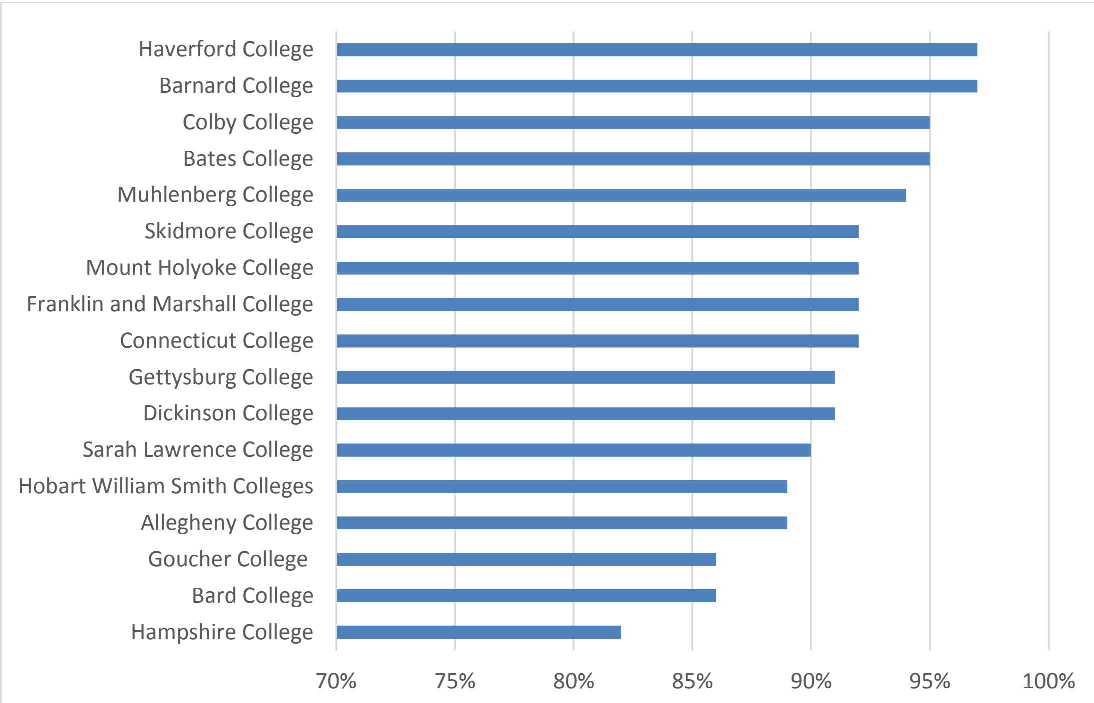
- In the book *How College Affects Students*, after reviewing approximately 2,500 studies on college students from the 1990s, and more than 2,600 studies from 1970 to 1990, the authors concluded that student engagement is a central component of student learning and success.
- Tinto (1993) identifies three major sources of student departure: academic difficulties, the inability of individuals to resolve their educational and occupational goals, and their failure to become or remain incorporated in the intellectual and social life of the institution.
- The national data suggests that historically marginalized student populations have received greater access to postsecondary education over the last decades. At Goucher College, 23% of first-time, degree-seeking students entering in Fall 2014 received the federal Pell grant and 9% are first-generation students (defined as neither parent received a Bachelor's degree). Access without support is not an opportunity.
- Retention has a significant impact on the college's budget similar to most small liberal arts colleges: tuition and fees as well as income from housing and dining have been the major revenue source (67%) for Goucher College. According to the most recent five years' financial data (FY 2011 through FY 2015), 46% of the College's total revenue was generated from undergraduate net tuition and fees and 21% from housing and dining income.

Goucher Data

- Goucher's fall 2013 to fall 2014 first-year retention rate is 77%, one of the lowest points in the institutional history. Ninety eight out of 401 students entering in Fall 2013 did not return for their sophomore year.



- With an 86% retention rate for the Fall 2012 cohort, Goucher's first-year retention rate is ranked 30th out of our 31 peer institutions.

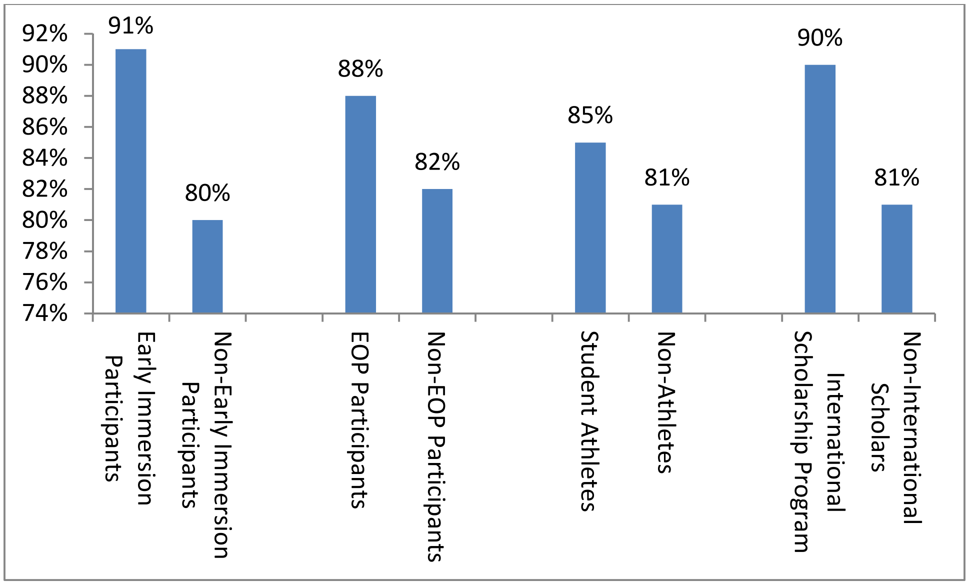


Data Source: IPEDS Fall 2012 cohort data (the most recent data available). Mid-Atlantic and New England peer colleges are presented in the chart.

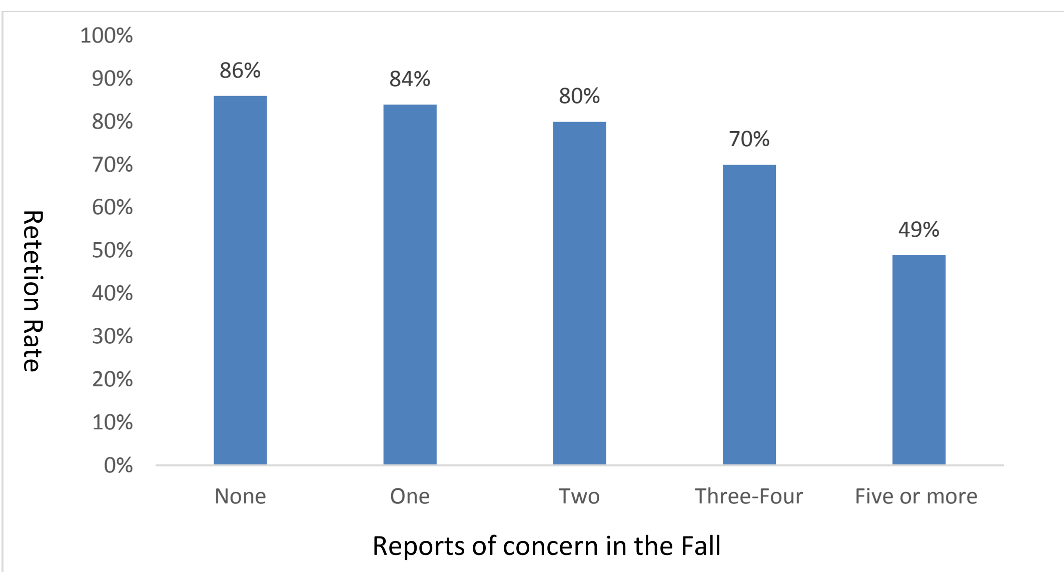
- Based on the most recent five cohorts' data, the strongest variable in predicting student retention at Goucher College is the first semester GPA.

	Coefficient	Standard Error	Statistical Significance	Odds Ratio
Fall GPA	**.421**	**.134**	**.002**	**1.523**
Age	**-.471**	**.167**	**.005**	**.624**
Early Immersion	**.544**	**.266**	**.041**	**1.724**
Student Athletes	**.418**	**.208**	**.044**	**1.519**
Total Fall Credits	**.162**	**.082**	**.047**	**1.176**
Number of Reports of Concern in the Fall	**-.191**	**.129**	**.050**	**.826**

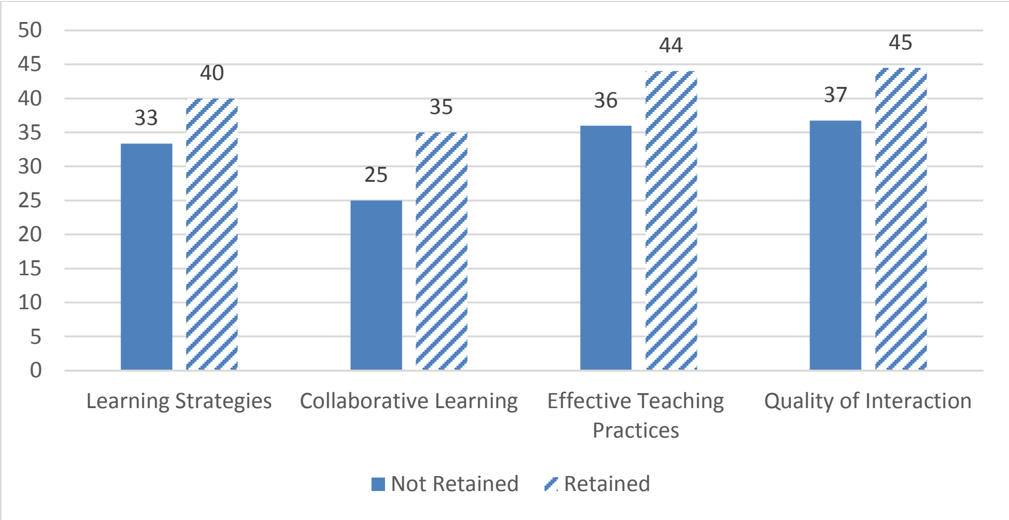
- At Goucher College, student athletes, Educational Opportunity Program (EOP) students, International Scholarship Program (ISP) students, and students who participated in the early immersion program are associated with a higher retention rate than their peers.



- A strong correlation exists between the number of concerns reported in the fall and student retention behaviors. Receiving three such reports in the fall indicates the critical point of severe alert. In the five-year dataset, 18 percent of Goucher's first-year students received three or more reports of concern in the fall semester.



- The Spring 2014 NSSE data suggests that compared to the students who did not return for their sophomore year, returning students reported a higher level of engagement, particularly on the indicators of collaborative learning, learning strategies, effective teaching practices, and quality of interaction.



Note: NSSE engagement indicator scores are calculated for each student and range from 0 to 60. The median scores of retained and not retained students are presented in the chart.

Questions:

1. Diagnose: In your view, what are the primary barriers for students who do not persist at Goucher College? Is the diagnostic information provided in the retention analysis aligned with your notion of the retention issue?

2. Design: Connecting with the retention analysis findings, what are the most promising institutional strategies and policies for overcoming those barriers? How can we collectively translate data into strategic actions?

3. Delivery: What role (s) do faculty and staff play in implementing the strategies and best practices to improve student retention rates at Goucher College?

Using Analysis to Drive Decisions in Improving Retention

Shuang Liu
Senior Director of Institutional Effectiveness
Goucher College
42nd NEAIR Annual Conference

Presentation Overview

- Goucher College
- Background
- Methodology
- Results
- Implications
- Turn Insights into Actions

Goucher College

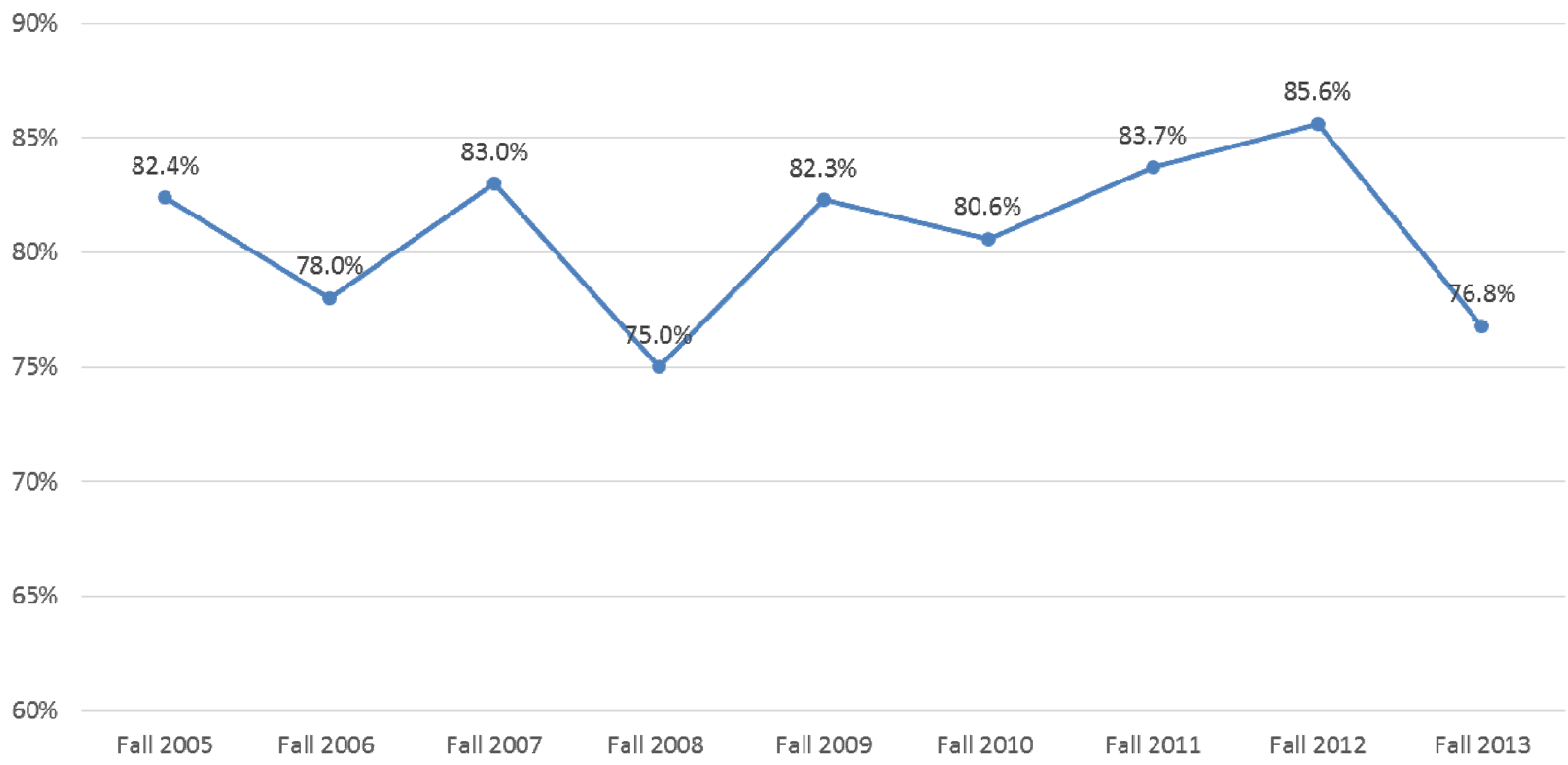
Goucher College is dedicated to a liberal arts education that prepares students within a broad, humane perspective for a life of inquiry, creativity, and critical and analytical thinking.

Goucher College

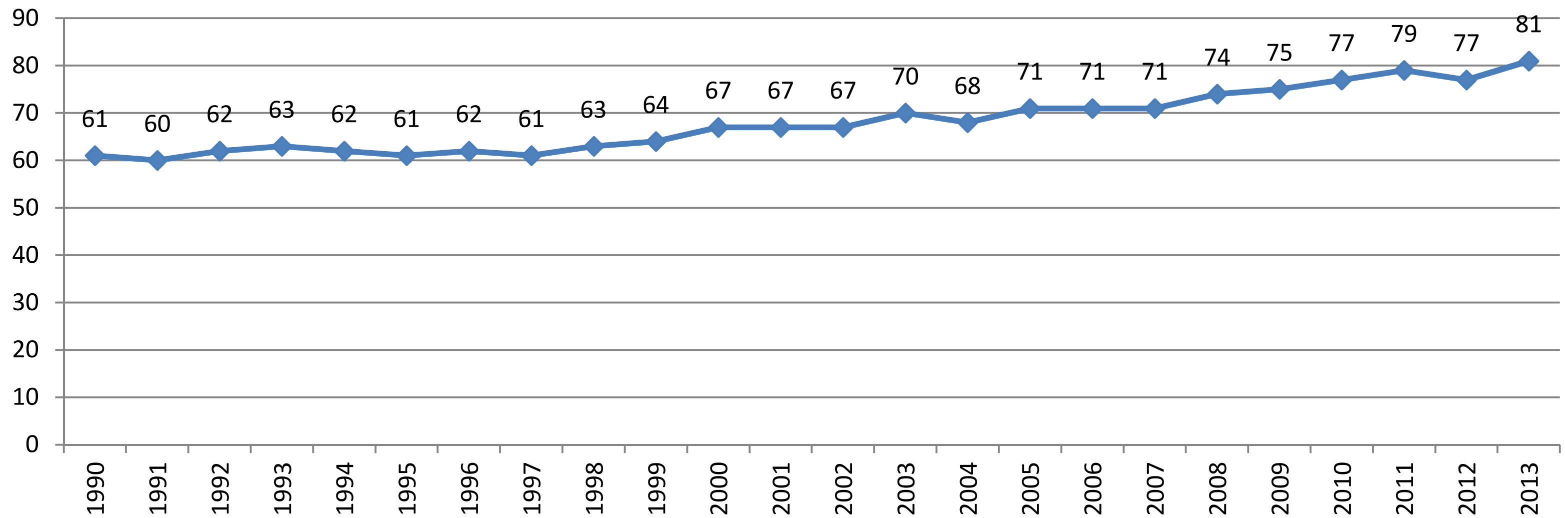
- Top 10 “Most Innovative School” (U.S. News)
- No. 1 in global education (U.S. News and others)
- One of 40 selected “Colleges That Changes Lives”
- GVA – first college to introduce alternative video application
- Undergrads from 44 states, 39 countries
- 10:1 student-to-faculty ratio
- 96% recent alums are employed or in graduate school

Challenges

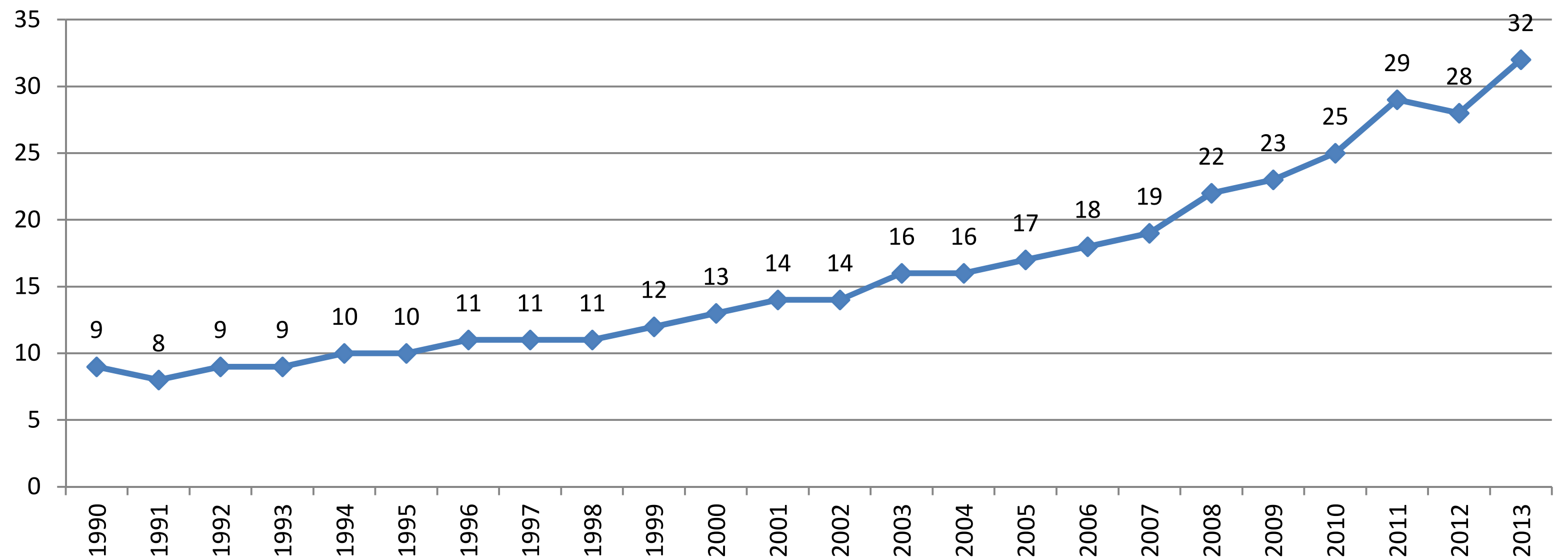
First-year Retention Trend Data



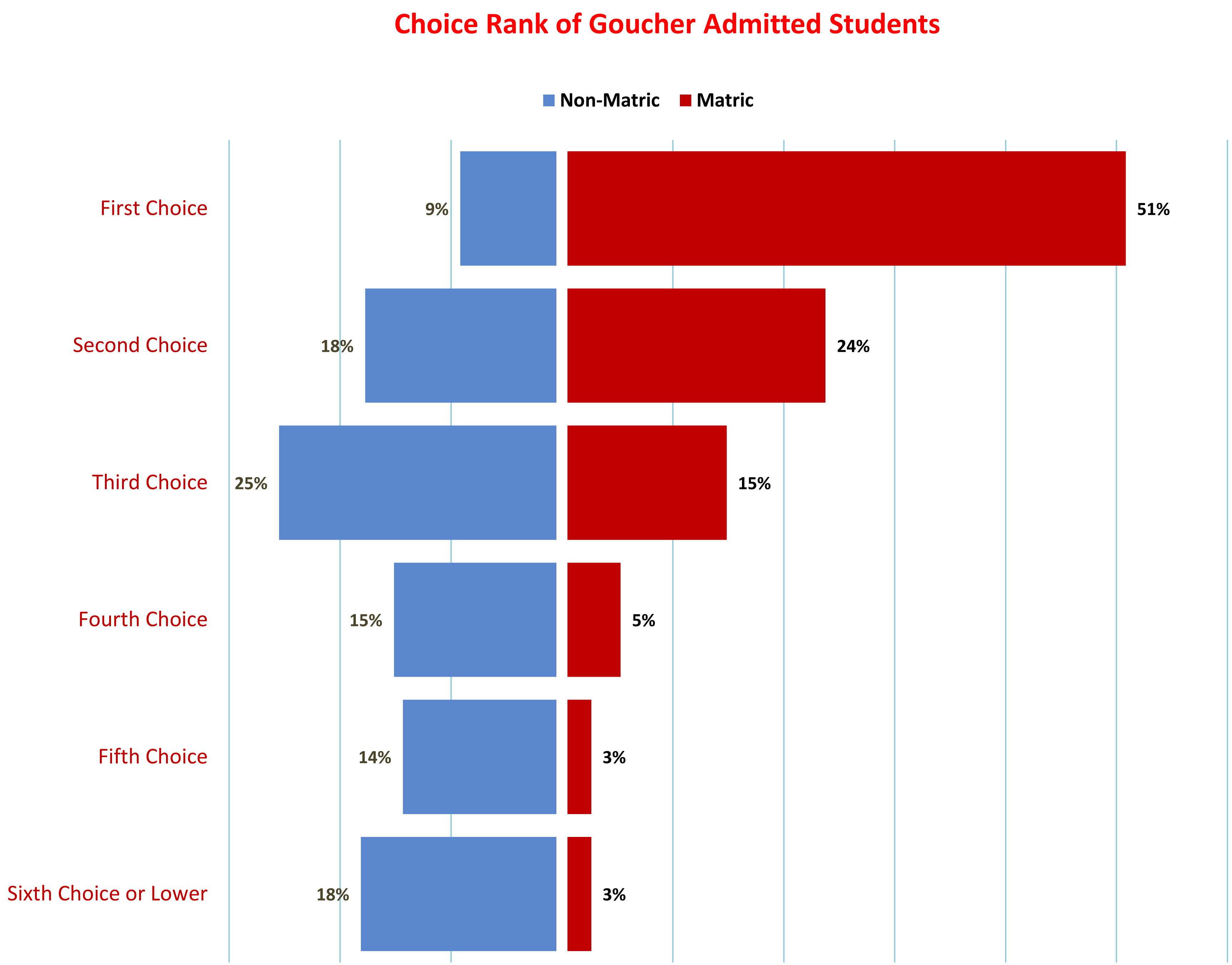
Percentage of Students Submitting Three or More College Applications



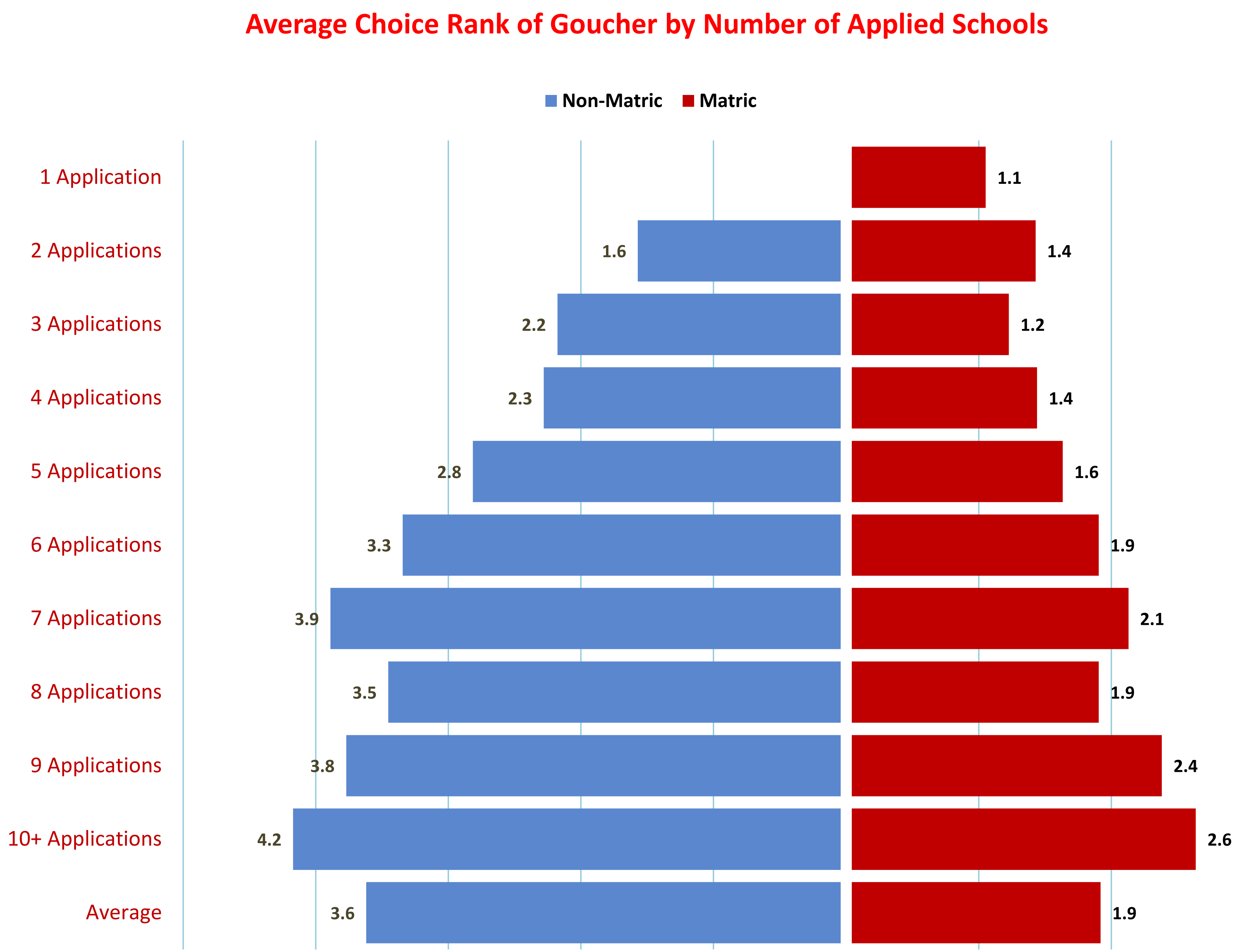
Percentage of Students Submitting Seven or More College Applications



Sources: Egan, K., Lozano, J.B., Hurtado, S., Case, M.H. (2013). The American Freshman: National Norms for Fall 2013. Los Angeles: Higher Education Research Institute, UCLA. Pryor, J.H., Eagan, K., Blake, L.P., Hurtado, S. Berdan, J., Case, M.H. (2012). The American Freshman: National Norms Fall 2012. Los Angeles: Higher Education Research Institute, Pryor, J.H., DeAngelo, L., Blake, L.P., Hurtado, S., Tran, S. {2007-211}. The American Freshman: National Norms for Fall. Report years 2007-2011. Los Angeles: Higher Education Research Pryor, J.H., Hurtado, S., Saena, V.B., Santos, J.L., Korn, W.S. (2006). The American Freshman: Forty Year Trends. Los Angeles: Higher Education Research Institute, UCLA.

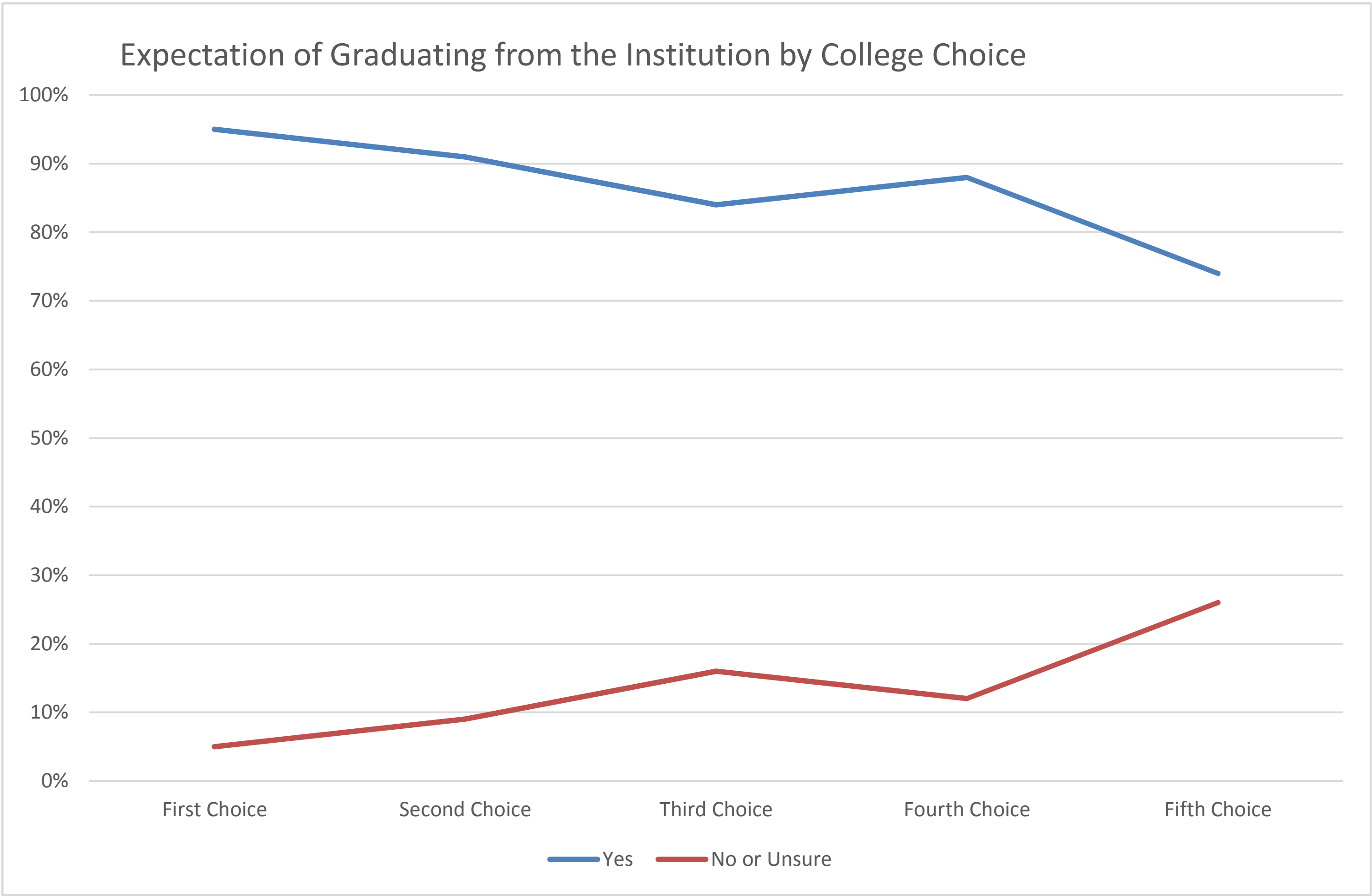


Source: 2015, HCRC Admitted Student Survey



Source: 2015, HCRC Admitted Student Survey

Challenges



Data Source: 2014 Beginning College Survey of Student Engagement Results

Methodology

- Research Question: What factors contributed to first-year retention?
- Data Source: Fall 2009-2013 cohort data (1,969 records)
- Dependent Variables: retained = 1, not retained = 0
- Independent Variables: 22 variables including demographics, incoming academic abilities, college engagement, etc.
- Logistic regressions were estimated to account for the predictive relationship between the independent and dependent variables.
- The retention indicator was merged with Fall 2013 BCSSE data.
- The retention indicator was merged with Spring 2014 NSSE data.

Methodology

- List of independent variables

Gender	EOP
Age	Resident
In-State/Out-of-State	Disability
Race	Nconcern Fall
Race-Asian	NAPR Fall
Race-African	Total Credit Fall
Race-Latino	Prcnt Fulltime Fall
Race-AmericanIndian	Fall GPA
Age	Spr GPA
Legacy Status	Year 1 Cum GPA
EFC	Dorm Plan
Family Income	Hourse1
SAT V	Hourse2
SAT M	Hourse3
SAT_Total	Hourse4
SAT W	Hourse5
ACT Read	Early E
ACT Eng	Athlete
ACT Sci	Retnew
ACT Math	Cohort
ACT Comp	International Scholar
HSGPA	
HS Name	
Writing Placement	
Math Placement	
Test Opt	
Admit Type	
Date Admitted	
Date Accepted	
NDays	

Results

	B	S.E.	Sig.	Exp(B)
Fall GPA	.421	.134	.002	1.523
Age	-.471	.167	.005	.624
Early Immersion	.544	.266	.041	1.724
Student Athletes	.418	.208	.044	1.519
Total Fall Credits	.162	.082	.047	1.176
Number of Reports of Concern in the Fall	-.191	.129	.050	.826

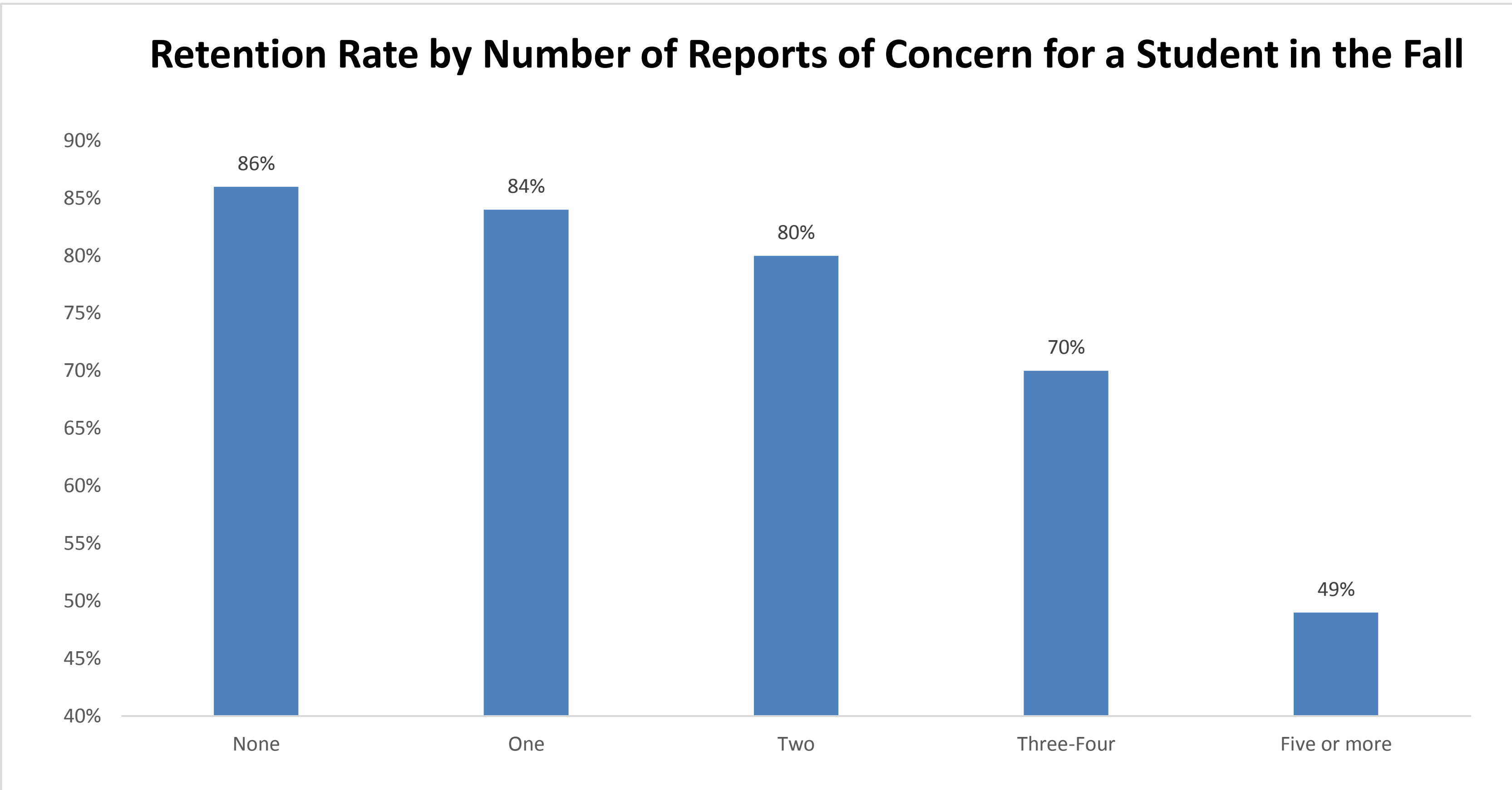
Communicating the Results

Data Brief 1: Focus on the logistic regression analysis

1.5x	Fall GPA	For every one point increase in Fall GPA, the odds of a student returning to Goucher for their sophomore year increased 1.5 times.
1.7x	Early Immersion	Students who participated in the early immersion program were 1.7 times more likely to return than their peers.
1.5x	Student Athletes	Student athletes were 1.5 times more likely to return to Goucher than non-athletes.
1.2x	Total Fall Credits	The number of credits students took in the Fall was positively related to retention. For every one credit increase, the likelihood of returning increased by 1.2 times.
0.6x	Age	Student age was found to be negatively associated with retention. Older students were more vulnerable to attrition. Starting at age 19, for every one year increase in age, the odds of returning decreased by 1.7 times (1/0.6).
0.8x	Number of Reports of Concern in the Fall	Students who received reports of concerns in the Fall were more likely to leave Goucher after their first year. For each report, the odds of a student returning decreased by 1.3 times (1/0.8).

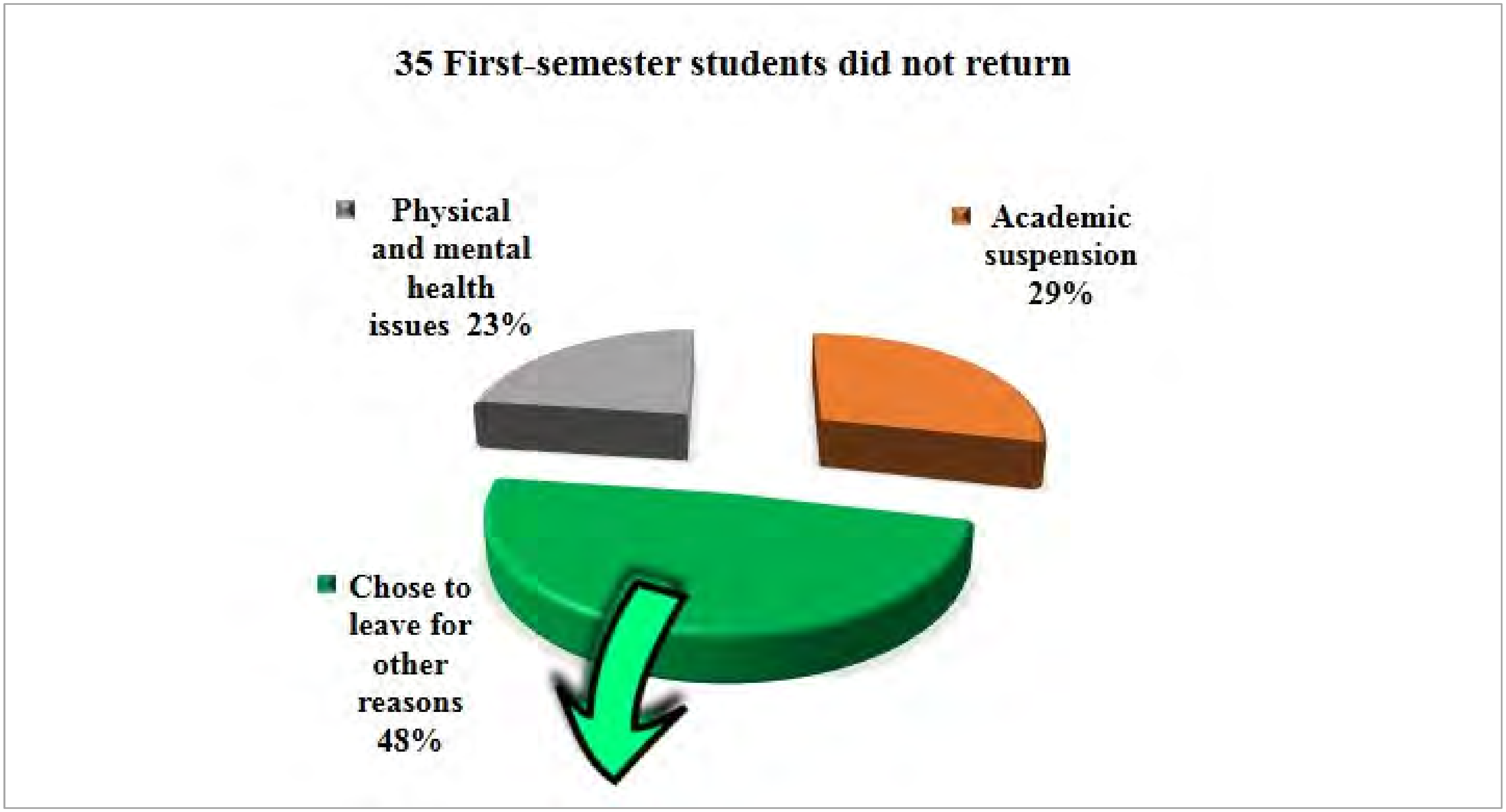
Communicating the Results

Data Brief 1: Focus on the logistic regression analysis



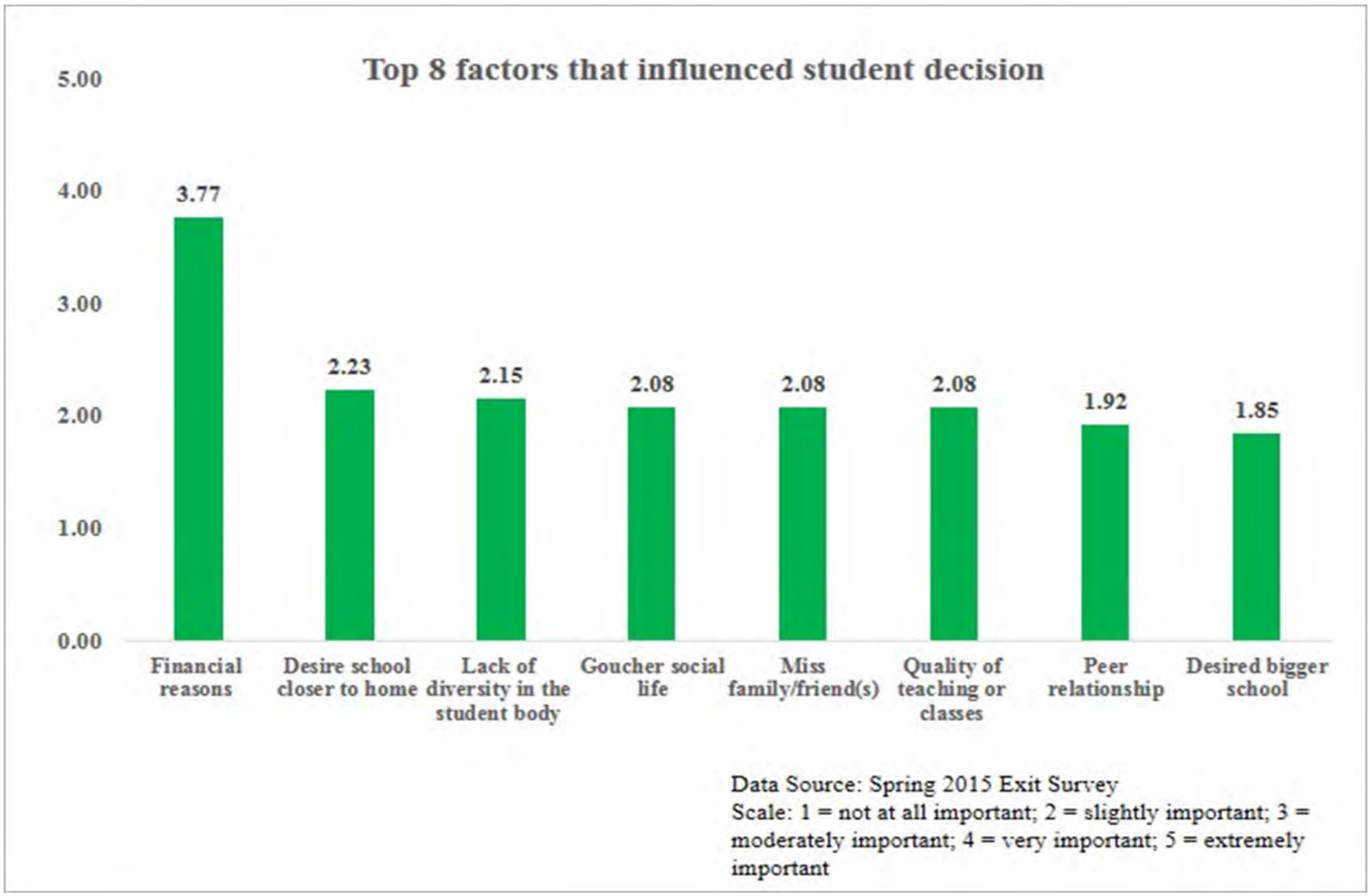
Communicating the Results

Data brief 2: Focus on the Fall 2014 to Spring 2015 retention data.



Communicating the Results

Data brief 2: Focus on the Fall 2014 to Spring 2015 retention data.



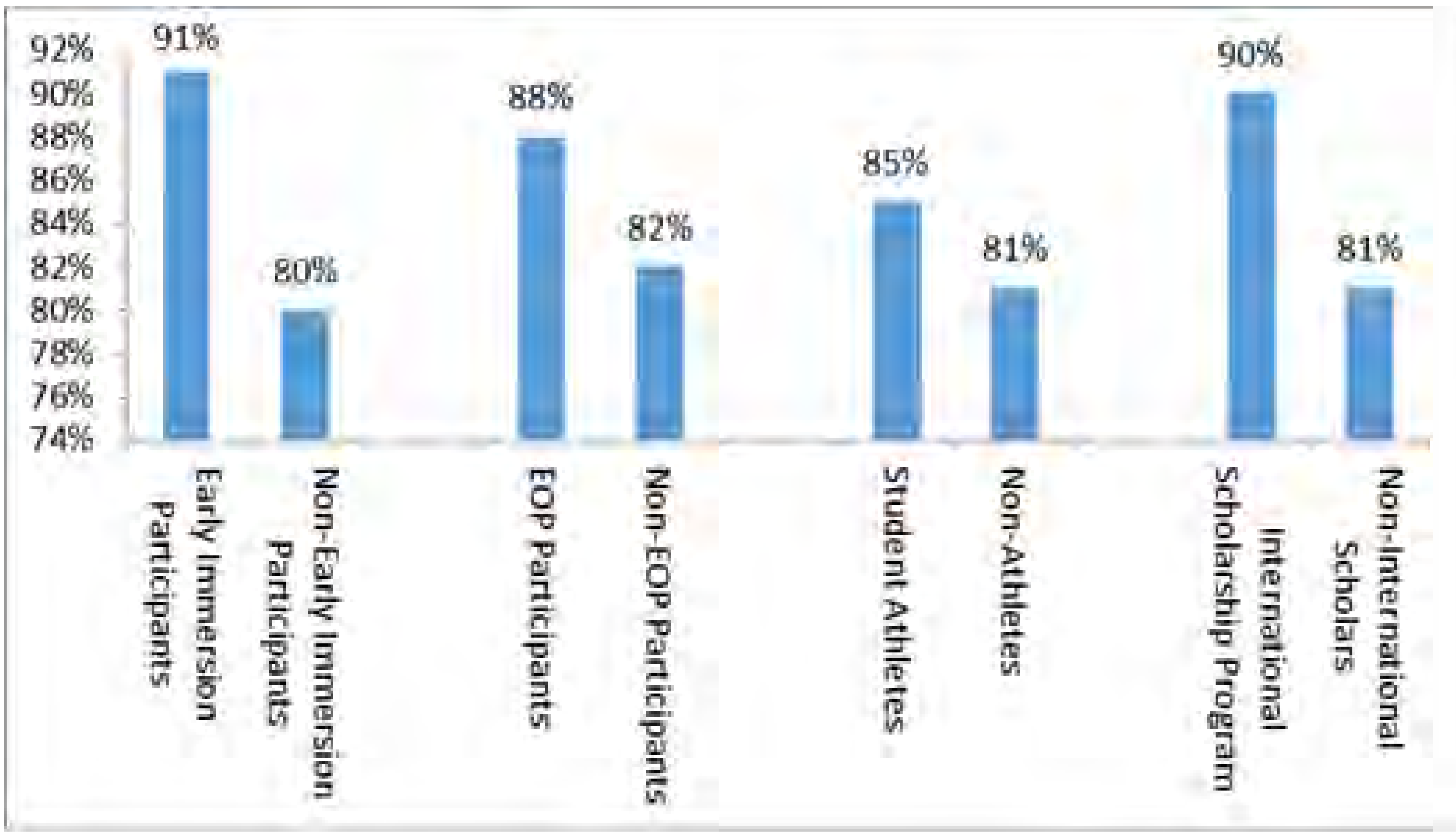
Communicating the Results

Data brief 2: Focus on the Fall 2014 to Spring 2015 retention data.

Subsequent Enrollment of Fall 2014 First-semester Non-returning Students			
Anne Arundel Community College			
Community College of Baltimore City			
Montgomery College - Takoma			
Salisbury University			
Towson University			
Colorado State University			
Community College of Vermont			
Delaware County Community College			
Kent State University			
Lake Forest College			
Louisiana State University			
N. Virginia Community College			
Norwalk Community College			
Portland State University			
SUNY Hudson Valley CC			
SUNY Westchester			
Temple University			
University of Delaware			
York College			

Communicating the Results

Data brief 3: Focus on the concept of student engagement



Communicating the Results

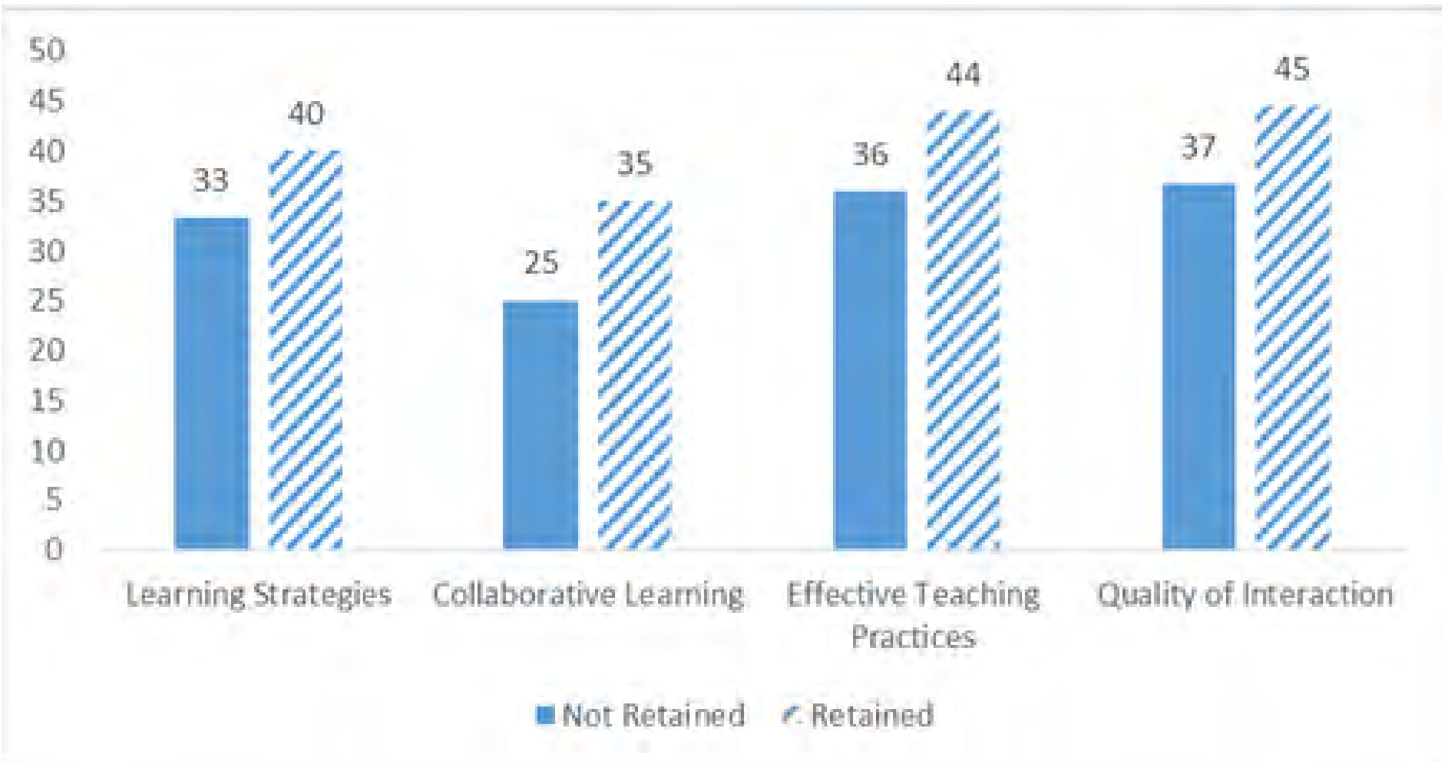
Data Brief 3: Focus on the concept of student engagement

NSSE Engagement Indicators		N	Mean	Std. Deviation	Std. Error Mean	Median
High-order Learning	Not Retained	9	36.11	12.94	4.31	35.00
	Retained	96	41.72	12.73	1.30	40.00
Reflective and Integrative Learning	Not Retained	9	30.16	11.17	3.72	34.29
	Retained	102	40.99	11.77	1.17	40.00
Learning Strategies	Not Retained	8	33.33	7.13	2.52	33.33
	Retained	92	41.88	13.42	1.40	40.00
Quantitative Reasoning	Not Retained	9	25.93	22.22	7.41	20.00
	Retained	100	25.47	14.54	1.45	23.33
Collaborative Learning	Not Retained	11	26.82	11.89	3.58	25.00
	Retained	106	36.42	12.42	1.21	35.00
Discussion with Diverse Others	Not Retained	8	51.88	8.84	3.13	52.50
	Retained	95	42.79	13.58	1.39	40.00
Student-Faculty Interaction	Not Retained	9	21.11	12.19	4.06	20.00
	Retained	98	22.60	14.43	1.46	20.00
Effective Teaching Practices	Not Retained	10	38.60	13.79	4.36	36.00
	Retained	99	43.54	10.01	1.01	44.00
Quality of Interaction	Not Retained	8	38.31	7.71	2.73	36.75
	Retained	96	43.96	9.05	0.92	44.50
Support Environment	Not Retained	8	37.19	15.44	5.46	36.25
	Retained	94	38.43	11.53	1.19	38.75

Data Source: Spring 14 NSSE Data: Engagement Scores by Retained and Not Retained Students

Communicating the Results

Data Brief 3: Focus on the concept of student engagement



Data Source: Spring 14 NSSE Data: Engagement Scores by Retained and Not Retained Students

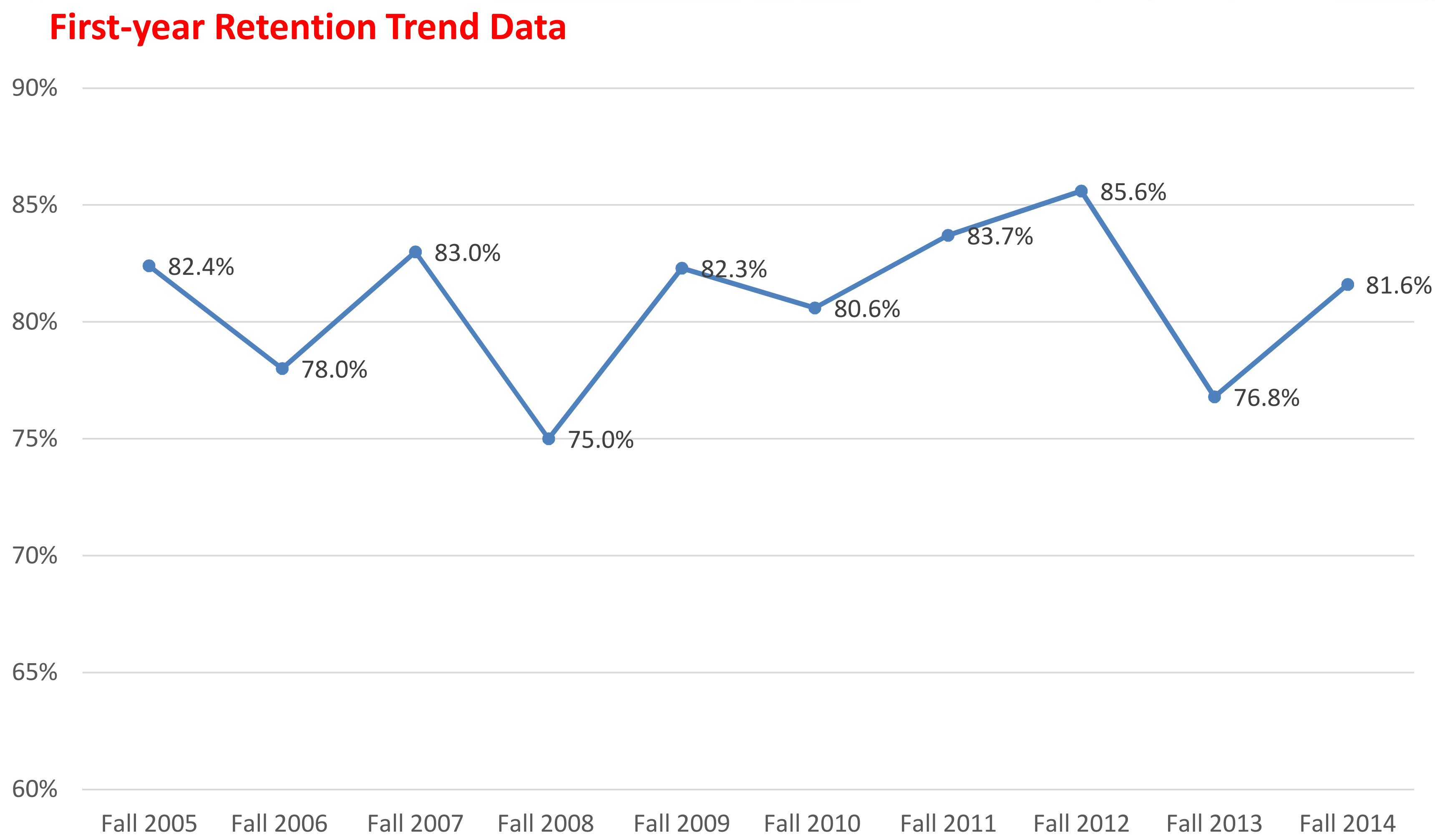
Implications

- APRs of concern and Fall GPA are the most significant predictors of first-year retention.
- Earlier APRs and graded work increases the effectiveness of student support tactics for all students.
- It is CRITICAL to track student access to support services.
- Student engagement indicators should be further collected.

Turn Insight into Action

- The Task force became a standing committee.
- Starfish was implemented to better track and identify at-risk students.
- APRs were replaced by Academic Progress Surveys in Starfish.
- BCSSE was added to the academic advising tool box.
- Academic suspension policy was revised.
- Academic probation policy was instituted.
- All students placed on academic probation were required to meet with an academic coach in ACE.
- Academic contract for student success contained both required academic activities and personalized academic goals.

Outcome



First-year Retention Trend Data
90%
85%
80%
75%
70%
65%
60%
82.4%
78.0%
83.0%
75.0%
82.3%
80.6%
83.7%
85.6%
76.8%
81.6%
Fall 2005
Fall 2006
Fall 2007
Fall 2008
Fall 2009
Fall 2010
Fall 2011
Fall 2012
Fall 2013
Fall 2014

The End

- Executive Summary
- Retention Data Briefs
- Faculty Retreat Round Table Summary
- Questions?
- Thank you!