LEADING THE CHARGE FOR INSTITUTIONAL RENEWAL

NEAIR 38th ANNUAL CONFERENCE

December 3–6, 2011

The Boston Park Plaza Hotel, Boston, Massachusetts

BOSTON CONFERENCE TEAM

President:
GAYLE FINK

Program Chair:
NICOLE MARANO

Associate Program Chair:
CHAD MUNTZ

Pre-Conf Workshop Coordinator:
MARK PALLADINO

Evaluation Coordinator:
LAURA UERLING

Newcomer/Mentor Coordinator:
MARY GOODHUE LYNCH

Poster Session Coordinator:
PAULA MAAS

Proposal Peer Review Coordinator:
MARK ECKSTEIN

Publications Coordinator:
CRISTI CARSON

Local Arrangements Chair:
MELANIE LARSON

Exhibitor Coordinator:
GURVINDER KHANEJA

AV Coordinator:
CHAD MUNTZ

Local Arrangements Coordinators:
ELIZABETH AVERY
DORIS CHOW-HANNON
MATTHEW HENDRICKSON
HEATHER ROSCOE

Conference Website Coordinator:
CHRISTOPHER CHONCEK

Administrative Coordinator:
BETH SIMPSON

NEAIR OFFICERS & STEERING COMMITTEE

Gayle Fink
PRESIDENT

Stephen Thorpe
PRESIDENT – ELECT

Heather Kelly
PAST PRESIDENT

Nancy Ludwig
RECORDING SECRETARY

George Rezendes
TREASURER

Ellen Boylan
MEMBER-AT-LARGE '11

Michael Dooris
MEMBER-AT-LARGE '11

Emily Dibble
MEMBER-AT-LARGE '12

Paula Maas
MEMBER-AT-LARGE '12

Mary Goodhue Lynch
MEMBER-AT-LARGE '13

Ann Marie Senior
MEMBER-AT-LARGE '13

Chad May
TECHNOLOGY CHAIR

Nicole Marano
PROGRAM CHAIR

Melanie Larson
LOCAL ARRANGEMENTS CHAIR

Dear NEAIR Colleagues,

The 38th Annual NEAIR Conference held in Boston, Massachusetts, December 3-6, 2011 encouraged attendees to contribute to *Leading the Charge for Institutional Renewal*. Three hundred and seventy-nine conference attendees had the opportunity to share and gain invaluable information from institutional research and higher education colleagues. The 2011 Conference Proceedings is a result of the conference theme in action.

The Conference Program team led by Program Chair **Nicole Marano** and Associate Program Chair **Chad Muntz** developed a program filled with plenty of variety that included three plenary/keynote sessions, eleven contributed papers, 26 workshares, twelve techshares, twelve special interest groups, and five table topics. Poster Session Coordinator **Paula Maas** organized sixteen posters to be on display. These offerings went through a blind peer review process facilitated by 59 proposal reviewers coordinated by **Mark Eckstein**. Pre-Conference Workshop Coordinator **Mark Palladino** organized fifteen workshops with 218 participants. Exhibitor Coordinator **Gurvinder Khaneja and Beth Simpson** partnered with a record eighteen exhibitors who offered seven exhibitor showcases and Lightning Talks.

Big thanks goes to Publications Coordinator **Cristi Carson** for all her hard work and keen eye editing the conference program, as well as compiling and organizing the 2011 Conference Proceedings. The 2011 Conference Proceedings contains papers submitted by authors, as well as the 2011 Best Paper Award recipients. The award recipients were determined by Best Paper Chair **Matthew Hendrickson** and his committee. The 2011 Best Paper this year is awarded to Leslie Stratton, Ph.D. and James Wetzel, Ph.D. of Virginia Commonwealth University for the paper, *Are Students Dropping Out or Dragging Out the College Experience? The Roles of Socioeconomic Status and Academic Background.* No Best First Paper was awarded from the Boston conference, and unfortunately, no submissions were received for the Boston conference Best IR Report/Practitioner award.

Local Arrangements Chair **Melanie Larson** and Local Arrangements Coordinators **Elizabeth Avery, Doris Chow-Hannon, Matthew Hendrickson, and Heather Roscoe** worked hard coordinating hotel, travel logistics and made sure we all enjoyed the *NEAIR Third Place* and all that Boston had to offer. AV Coordinator **Chad Muntz** assisted with technology and Dine-Around Coordinator **Doris Chow** made sure we were well-fed and had an additional networking opportunity.

Conference Website Coordinator **Chris Choncek** and Administrative Coordinator **Beth Simpson** developed and maintained the conference website as well as conference registration. Next year's conference planning will be facilitated by online evaluations analyzed by Evaluation Coordinator **Laura Uerling**.

It was a pleasure to work with such an extraordinary Conference Planning Team and the many talented volunteers. A premiere professional development opportunity was the result of the efforts of these individuals. We hope you take advantage of all the great information the 2011 Conference Proceedings have to offer!

Wishing you all the best,

Gayle Fink
NEAIR President 2010-2011

Leadership & Committees 2010 – 2011

NEAIR
North East Association for Institutional Research

Officers:	President	Gayle Fink
	President-Elect	Steve Thorpe
	Recording Secretary ('08-'11)	Nancy Ludwig
	Treasurer ('11-'14)	George Rezendes
Steering Committee Members:		
	Past President	Heather A. Kelly
	Program Chair	Nicole Marano
	Local Arrangements Chair	Melanie Larson
	Technology Chair	Chad May
	Member-At-Large ('08-'11)	Ellen Boylan
	Member-At-Large ('08-'11)	Michael Dooris
	Member-At-Large ('09-'12)	Emily Dibble
	Member-At-Large ('09-'12)	Paula Maas
	Member-At-Large ('10-'13)	Ann Marie Senior
	Member-At-Large ('10-'13)	Mary Goodhue Lynch
Administrative Coordinator (ex-officio)		Beth Simpson

Standing Committees

Local Arrangements (Standing Committee)

Chair	**Melanie Larson**
Exhibitor Coordinator	Gurvinder Khaneja
AV Coordinator	Chad Muntz
LAC Coordinators:	Elizabeth Avery
	Doris Chow-Hannon
	Matthew Hendrickson
	Heather Roscoe

Membership (Standing Committee)

Chair	**Mary Goodhue Lynch**
Chair-elect	TBD
Member	Laura Uerling
Past Chair (ex officio)	Paula Maas

Nominations (Standing Committee)

Chair	**Heather Kelly**
Member – 4 YR Private Sector	Jane Kimble
Member	Kelli Armstrong
Member	Rommel Guadalupe
Member	Chad Muntz
Member	Denise Nadasen
Member	C. Ellen Peters

Prof Development Services (Standing Committee)

Chair	**Steve Thorpe**
Member - Program Chair-Elect	Mark Palladino
Member	Michelle Appel
Member	Ellen Boylan
Member	Paula Maas
Member	Linda Mallory

Program (Standing Committee)

Chair	**Nicole Marano**
Associate Program Chair	Chad Muntz
PCW Coordinator	Mark Palladino
Evaluation Coordinator	Laura Uerling
Mentor Coordinator	Mary Goodhue Lynch
Newcomer Coordinator	Paula Maas
Peer Review Coordinator	Mark Eckstein
Poster Session Coord.	Paula Maas
Publications Coordinator	Cristi Carson

Site Selection (Standing Committee)

Chair	**Steve Thorpe**
Treasurer	George Rezendes
Member	Tracey Barnes
Member	Michael Dooris
Member	Jessica Shedd
Admin coord (ex officio)	Beth Simpson

Technology (Standing Committee)

Chair	**Chad May**
Conf Web Coordinator	Chris Choncek
Member	Dan Nugent
LAC (ex officio)	Melanie Larson
Program Chair (ex officio)	Nicole Marano
Admin Coord. (ex officio)	Beth Simpson

NEAIR
North East Association for Institutional Research

Leadership & Committees 2010 – 2011

Ad Hoc Committees

Finance Committee

Chair	**Emily Dibble**
Treasurer	George Rezendes
Member	Rebecca Brodigan
Member	Alan Sturtz
Past Treasurer	Eleanor Swanson

Grants Committee

Chair	**Michael Dooris**
Chair-Elect	Ann Marie Senior
Member	Cathy Alvord
Member	Diane Petruccio
Member	Ingrid Skadberg
Past Chair (ex officio)	George Rezendes

Incorporation Committee

Chair	**Heather Kelly**
Member	Michael D'Alessandro
Member	Mitchell Nesler
Treasurer	George Rezendes
Member	Jessica Shedd
Member	Sean Simone
Past Treasurer	Eleanor Swanson
Admin Coord. (ex officio)	Beth Simpson

Conference Proposal Peer Reviewers

Mark Eckstein, Peer Review Coordinator

Henry C Alphin Jr
Louis Attinasi
Meredith Billings
Felice Billups
Jason Canales
Betsy Carroll
Cristi Carson
Marlene Clapp
Elizabeth Clune-Kneuer
Theodore Coladarci
Lisa Daniels
Cherry Danielson
Monica Defendorf
Emily Dibble
Suhua Dong
Karen Durkin
Nora Galambos
Cheryl Goldstein
Joan Graham
Jennifer Gray
Phil Handwerk
Matthew Hendrickson
E. Jill Hirt
May Hser
Elizabeth Johnston-O'Connor
Lisa Keating
Ethan Kolek
Arthur Kramer
Marco Krcatovich
Elisabeth Lackner
Pamela Lamborne
Gregory LaPointe
Shannon Tinney Lichtinger
Linda Mallory
Chad May
Pat Mizak
Dale Mort
Kevin Murphy
Mitchell Nesler
Doug Nutter
Tiffany Parker
Barton Pursel
Cheryl Rollins
Heather Roscoe
Timothy Rupright
Stephen Sheridan
Yan Shi
Sri Sitharaman
Jacquelyn Stirn
Alan Sturtz
Kristen Sullivan
Bruce Szelest
Steve Thorpe
Laura Uerling
Carol VanZile-Tamsen
Lynn Wallace
Michael Whitcomb
Alexander Yin
Lillian Zhu

Table of Contents

*Winner 2011 Best Paper Award

ACKNOWLEDGMENT

Contained within these pages of the NEAIR 38th Annual Conference Proceedings are eight contributed conference papers/presentations authored by thirteen NEAIR colleagues.

Additional conference presentations are just a few clicks away—accessible within the NEAIR website under the Conference Tab. These pages are only accessible to signed in NEAIR members.

Special thanks go out to Gayle Fink, Nicole Marano, and Beth Simpson for their contributions, oversight and support with all aspects of publication responsibilities during the course of this past year.

Cristi Carson
2011-2012 NEAIR Publications Chair
University of Southern Maine

JEL: I24, I2

Are Students Dropping Out or Dragging Out the College Experience?

The Roles of Socioeconomic Status and Academic Background

Leslie S. Stratton

and

James N. Wetzel[*]

RRH: STRATTON & WETZEL: DROPPING OUT OR DRAGGING OUT?

September 2011

JEL: I24, I2

Are Students Dropping Out or Dragging Out the College Experience?

The Roles of Socioeconomic Status and Academic Background

ABSTRACT

Disadvantaged students are substantially less likely to complete a college degree in six years than more advantaged students. The majority of the race/ethnicity differential and 20-35% of the family income and parental education differential is explained by academic background. However, 36% of those without a degree are still enrolled. When taking such persistence into account, we find Hispanics are less likely to have graduated because they are more likely to drag out the college experience, not because they have dropped out. On the other hand, first generation college students appear to be at greater risk of dropping out, rather than persisting.

Leslie S. Stratton
Professor
Department of Economics
Virginia Commonwealth University
IZA Bonn, Germany

James N. Wetzel
Professor
Department of Economics
Virginia Commonwealth University

Are Students Dropping Out or Dragging Out the College Experience?

The Roles of Socioeconomic Status and Academic Background

I. INTRODUCTION

Substantial differences in achievement by socioeconomic status have been documented throughout the educational structure from K-12 through higher education and are a frequent subject of discussion in the public policy arena. The No Child Left Behind Act of 2001 places substantial pressure on K-12 educators to eliminate such differences. College admissions offices have been encouraged to support equal opportunity/access policies for decades. More recently, colleges are being pushed to improve student outcomes, with public financial support becoming contingent on performance. However, it is critical to control for academic background in order to distinguish the differential impact of socioeconomic status as compared to academic background on college outcomes. If academic background is important, then policy makers could focus further attention and resources on K-12 education in order to improve not only K-12 education itself but also college outcomes. If socioeconomic status remains important after controlling for academic background, then policy changes aimed at supporting those at risk at the college level, as identified by socioeconomic status, may be necessary to increase college success.

Furthermore, the traditional approach to measuring 'success' in college by relying only on graduation rates may be misleading. Graduation rates vary depending upon the time period of the analysis. Data that were once assessed using four-year graduation rates as the goal are now commonly assessed using six-year graduation rates, but even this extended measure fails to capture the substantial persistence observed amongst those who have not graduated. We find that 36% of those who have not graduated at the end of six years are still enrolled. These students are not necessarily "failures"; they may simply be taking longer to graduate. It is important from both a research and a public policy perspective that statistical analysis take into consideration not only degree receipt but also enrollment status when last observed.

We perform such a statistical analysis using the 1996-2001 Beginning Postsecondary Survey. These data comprise a national sample of undergraduates whose enrollment status is observed for six years even as they move from one institution to another. Restricting the analysis to those initially enrolled at four-year institutions, we find that controlling for academic preparation/ability substantially reduces the gap in graduation rates between less and more advantaged socioeconomic groups, particularly for African Americans and somewhat less so for first generation college students. There remains a significant six to 11 percentage point differential in graduation rates for students from lower income and less educated households. More importantly, we also find that those who are still enrolled six years following matriculation are substantially different from both those who are no longer enrolled as well as those who have graduated, and that the marginal impact of socioeconomic status on persistence differs across the population. For example, being Hispanic is associated with greater persistence, whereas being a first generation college student is associated with a higher probability of non-enrollment at the six year mark.

II. LITERATURE REVIEW

A substantial body of research addresses the decision to attend college. Much of it is based on Becker's (1964) model of education as an investment in human capital. According to this theory, individuals pursue a college degree if the expected net present value associated with doing so is positive. If one focuses narrowly on financial aspects, the benefits are the increased financial earnings of a college graduate relative to those of a high school graduate and the costs are the direct costs such as tuition and books as well as the indirect costs in the form of foregone earnings while in college. Taking a broader perspective, benefits include the various psychic and social benefits associated with college attendance and costs include the time away from family responsibilities as well as the sacrifice of leisure time to class attendance and to study time.

Initial enrollment differences, also referred to as "access", by socioeconomic status, have declined over the last decades but remain substantial. Socioeconomic status is captured here by race/ethnicity, family income, and parental education. The College Board (2010) reported that in 2007 on average 67.2% of high school graduates were enrolled in a two- or four-year college immediately after graduation. The comparable figure for Hispanics was 60.9%, for African Americans 55.6%, for those from low income households 55.0%, and for those whose parents had completed no more than high school 50.9%. If some student subgroup, such as Hispanics, is "under-represented," the converse subgroup (non-Hispanics) is necessarily overrepresented relative to the population average.

Differences by socioeconomic status would be consistent with theory if the costs and/or benefits varied systematically for these different populations. Barrow and Rouse (2005) find no differences in the returns to education by race/ethnicity. Kane (1994) explains aggregate trends in college enrollment for African Americans during the 1980's as a function of changes in college costs and parental background. He is unable to control for academic ability but posits that ability is closely linked to family background. Cameron and Heckman (2001) find substantial differences in school achievement at age 15 for men by race/ethnicity but, in further analysis, find that controlling for long-term family income, academic ability, and family background (including parental education) explains all these substantial differences and more. They conclude that African American and Hispanic men are actually more likely to complete high school and attend college than white men if these control factors are taken into account. Carneiro and Heckman (2002) find that the family income-college enrollment relation is primarily driven by pre-enrollment differences by family income in ability. Vignoles and Powdthavee (2009) report that income differences in college attendance in the UK are entirely explained by academic background. Belley and Lochner (2007), however, report that, particularly for the less able, income has become a more important factor in driving college enrollment in the U.S., and Dynarski (2003) documents a relation between financial aid and enrollment as well as completion that suggests a link to income. Clearly then it is important to include controls for all measures of disadvantage – socioeconomic and academic – in order to accurately gauge the importance of each factor individually. Unfortunately the data necessary to do so can be difficult to obtain, particularly when looking at the educational attainment of older individuals (Deming and Dynarski 2009).

Even if access to higher education were independent of socioeconomic status, educational attainment may not be. Enrollment in college does not guarantee graduation. The College Board (2010) reports that on average only 56.1% of those entering college, even with the intent of earning a bachelor's degree, graduate within six years. While these 2007 figures reflect an increase from 1997, they clearly demonstrate that degree attainment is far from universal.[1] Furthermore, for African Americans the comparable figure is 40.5%, while for Hispanics it is 46.8%. Recall that these measures are contingent upon enrollment and so are already reflective of a selective population.

While theory suggests that only students who expect the benefits of pursuing a college degree to exceed the costs enroll in college, this does not preclude dropout. Expectations may change as students obtain new information about their expected returns.[2] This new information could be relative to their academic ability, the cost of college, or their likely future returns. Students coming from less advantaged households may be more likely to update their expected net benefits as they may have had less accurate information when making the decision to enroll in the first place. As educational attainment is lower for these populations in general, their knowledge of the process and its net benefits is likely less accurate.

There is a substantial literature pertaining to graduation in the education field,[3] less so in economics. Often these studies employ data on students from only a single institution (see for example, DesJardin, Ahlburg, and McCall 1999), even though Adelman (2006) finds that as many as 60% of all undergraduates attend multiple institutions. Some notable exceptions include Adelman (2006) who uses NELS data, Cameron and Heckman (2001) who use NLSY data, and Cragg (2009) who uses the data employed here. Each demonstrates the importance of controlling for academic background. Adelman in particular argues that controls for test scores, high school grades, and high school curriculum are all important and jointly dominate the impact of socioeconomic status.

Graduation in these analyses is typically modeled as a binary outcome occurring within a fixed time frame. Those who have not graduated within that time frame are treated as a homogenous population. Work in the persistence literature suggests that this assumption may be unwarranted. In examining student persistence from the first to the second year of college, Stratton, O'Toole, Wetzel (2008) find significant differences within the population of non-persisters (those not enrolled one year following matriculation) between those who reenroll within the subsequent 12 months and those who do not. If all degree recipients completed their requirements within a fixed period of time, measuring success using only degree receipt would fully capture the variable of interest. However, students seem to be taking longer and longer to complete their requirements. Attention these days is focused on six-year graduation rates. These six-year numbers are the rates that four-year institutions are required to provide under the 1990 Student Right-to-Know Act. Even following students for only six years may not be sufficient to identify all 'successful' undergraduates. We address this censoring by using information on enrollment six years following matriculation to distinguish between individuals who are still enrolled in college (persisters) and those who are not (non-persisters), while also identifying degree recipients.

Focusing on populations that have been historically underrepresented at postsecondary institutions, we contribute to the literature (1) by expanding the set of six-year college outcomes to recognize not just those who have completed their degree, but also those who are still persisting in their studies, and (2) by using a representative national data sample of younger college students that follows students as they move between institutions and includes detailed information on respondents' test scores, high school grades, and high school curriculum.

III. ANALYSIS FRAMEWORK

Standard analyses of six-year college outcomes use a logit model to distinguish between those who graduate and those who do not. We begin by estimating such a simple logit controlling only for gender, race, ethnicity, parental education, household income, age, unemployment rate, and marital and parental status. We use these results to estimate the marginal impact of socioeconomic status as measured by race, ethnicity, parental education, and income on graduation probabilities. These marginal results tell us the impact of each factor, ceteris paribus. We then add controls for academic background/ability and recalculate the marginal impact of socioeconomic status to determine the degree to which socioeconomic status rather than academic preparedness influences graduation rates. Finally, we estimate a specification that controls for a broad array of additional covariates sometimes included in attrition and/or graduation studies to assess the impact these other controls have on observed marginal effects by socioeconomic status. These three steps mimic those employed by Vignoles and Powdthavee (2009) to analyze persistence in the UK.

We then take an important, additional step to call attention to the distinction between persistence and dropout. Specifically, we expand the traditional analysis to further distinguish between those who are enrolled in the last term and those who are not. This analysis requires estimation of a multinomial logit specification. The application is much like that in Stratton, O'Toole, and Wetzel (2008) who use a multinomial logit specification to distinguish between continued enrollment, stopout, and dropout in the first year of college. Thus, the same specifications estimated for the simple logit are rerun for the richer multinomial logit specification to calculate the marginal impact socioeconomic status has upon this three-fold and much more meaningful measure of college outcomes. This analysis will allow us to determine whether some less advantaged populations might have lower graduation rates, not because they are no longer engaged but rather because they are taking longer to graduate.

IV. DATA

The data employed in this analysis come from the restricted access 1996-2001 Beginning Postsecondary Survey (BPS) collected by the National Center for Educational Statistics (NCES) of the Department of Education. These data constitute a nationally representative sample of students who first matriculated to a postsecondary institution in the 1995-1996 academic year. We restrict our analysis to those individuals with enrollment information through spring 2001 so that we have adequate time to track progress. Given the focus on academic programs culminating in a Baccalaureate degree, enrollment at less than two-year institutions and other institutions which are not likely to offer academic credit (such as beauty, training, and trade schools) is ignored. Some of those initially attending a two-year school are seeking a Baccalaureate degree. However, due to the unobserved and heterogeneous goals of this population, we follow common practice and further restrict our analysis to those in the sample who initially enrolled at a four-year institution. Subsequent enrollment at a two-year institution is recognized. These restrictions yield a sample of 6190 individuals.

Information on academic preparation and student ability is critical for this analysis. These data are missing for a substantial fraction of older students and those not from the United States. As a result, students from abroad and students age 23 and above are excluded from the analysis. A handful of individuals are excluded due to missing age or other characteristics of interest. These restrictions leave a final estimation sample of about 5820 individuals.[4] Sample statistics for this population are reported in Table 1. All the results reported here utilize the BPS longitudinal weights so as to replicate a nationally representative sample; all statistical estimates are corrected for the BPS's complex survey design.

Detailed personal information is available for every respondent. This includes information on gender, race, ethnicity, and age; state of residence; marital and parental status; and parental education and income. State of residence is used to match the state's 1995 unemployment rate to the sample. Higher unemployment rates imply a lower opportunity cost associated with college enrollment and may attract a different population of students. Regional dummies are also incorporated. Parental education is identified based on the reported education of the most educated parent, with preference given to parental reports. College degree receipt is the modal response. Almost no student's parents were high school dropouts. We distinguish between those parents with no more than a high school degree, those with some college, and those with a post-graduate degree using dummy variables. First generation college students are variously defined in the literature as either those whose most educated parent has no more than a high school degree or those whose most educated parent has less than a college degree: our specification allows for either definition. A dummy variable is used to identify respondents who declare they are independent of their parents, and income dummies that approximately split the population into quartiles are employed to allow a non-linear income effect. The highest income quartile is treated as the base case.

Academic preparation/ability is captured using a number of different variables, as suggested by Adelman (2006). A dummy variable to indicate high school degree receipt is incorporated to identify graduation and perhaps the character trait 'persistence'. Less than 2% of our sample do not have a degree. A measure of the most advanced math course the student plans to take is included to capture the rigor of the student's high school curriculum. Approximately 11% of the sample fails to report this information. We use a dummy variable to identify these persons and treat Trigonometry as the base case. Alternative specifications using NCES coding of the quality of the student's high school curriculum yield substantially the same results. Standardized SAT test scores and self-reported high school GPA are used to assess individual ability. Again dummy variables are used to identify those with missing values. Students taking the ACT are identified with a dummy variable and their ACT scores converted to SAT scores using a concordance table published by the College Board (1999). Grades are self-reported, since high school transcripts were not available, and such reports are likely biased upward (more students report an A average than any other outcome). Each of these measures of academic preparation/ability is determined prior to college enrollment. As such this research avoids the endogeneity problem associated with using first year college grades to assess progress towards a degree.

In our final specification, we include information on a wide variety of other factors sometimes incorporated in studies of college outcomes. For example, information on the first institution attended is incorporated at this stage. Specifically, we include controls for institution type (public/private), size, growth rate, and institution selectivity. The Integrated Postsecondary Education Data System (IPEDS) from NCES was used to identify the type, size, and growth rate of the institution. Type and size are commonly included as covariates. The growth rate of the institution over the previous four years is included as a proxy for resource availability. Work by Bound, Lovenheim, and Turner (2010) suggests that students may have difficulty completing their studies at institutions experiencing exceptional enrollment growth. Barron's admissions competitiveness index ratings for 1992 were used to classify institution selectivity (Schmitt 2009). There is substantial evidence that more selective schools have higher success rates all else constant (see, for example, Cragg 2009).

Data on the receipt of financial aid in the first year is also included at this stage. We know which individuals received grants, loans, and/or work-study aid. There are concerns about the accuracy of the reported dollar values. The dollar values also have different implications for enrollment decisions at different institutions given the substantial variation in tuition rates across institutions, as tuition levels affect the unmet need that influences both the receipt of and the dollar amounts of financial aid. Thus, we follow Hu and St. John (2001) and Johnson (2008) in using dummy variables to take into account financial aid type. The modal respondent used as a base case received some grant aid.

Finally, a dummy variable to identify those who first enrolled in spring 1996 rather than fall 1995 is incorporated. Those not enrolling in fall 1995 may be more marginal students either from an institutional perspective or from a motivational perspective – a factor particularly important in Bean's (1980) model of attrition. Note that all of the variables added in the final specification could be considered endogenous. Institutional characteristics are effectively chosen by the student in deciding to enroll. Financial aid offers are also often institution-specific. Finally, the decision to attend college clearly encompasses the decision of when to attend. Endogenous covariates can bias parameter estimates. However, while such covariates are endogenous as regards the decision to attend itself, our sample is already conditional upon attendance. Given this, one might consider such covariates predetermined for the research issue we address. Thus, while focusing our discussion on our more restricted specifications, we also present results for this expanded specification to assess the sensitivity of our results to the inclusion of such covariates. Behaviors such as stopout and part-time enrollment delay graduation but also represent decisions students make along the way and hence are clearly endogenous with respect to six-year outcomes. To avoid such clear endogeneity, we never include controls for actions taken post-enrollment.

The outcome measures for our analysis are derived using information on Baccalaureate degree receipt and college enrollment at the conclusion of spring 2001. Mimicking previous studies of college outcomes, we construct a simple binary outcome measure to identify those individuals who have graduated as of spring 2001. Column 1 of Table 2 presents average graduation rates for each of the socioeconomic indicators used in this analysis. These measures are slightly higher than those generally reported as they capture graduation at any institution. The overall fraction of the sample that graduates is 63%. The fraction graduating from the first institution attended (not reported in the table) is 55% - a number that matches the six-year graduation rate calculated using IPEDS data for the 2006 cohort. We find evidence (available upon request) that less advantaged populations are more likely to attend multiple institutions, but no evidence that controlling for this alters the results reported below. Proceeding down Table 2, our sample graduation rates are slightly higher at 66% for whites, and substantially lower at 45% for African Americans and 54% for Hispanics. Graduation rates are lowest for those whose most educated parent has no more than a high school diploma (50%) and highest for those with a parent who has a post-graduate degree (77%). Finally, graduation rates rise from 50% for those with the lowest family income to 76% for those with family incomes of at least $75,000. Raw differences indicate a graduation rate differential of about 21 percentage points for African Americans (66%-45%), ten for Hispanics, 19 for those having the least educated versus college educated parents, and 25 for students from the lowest versus highest income quartiles.

We are also, however, able to distinguish between those who did not graduate but are still enrolled in spring 2001 (henceforth called 'persisters') and those who did not graduate and are not enrolled in spring 2001 (henceforth called the 'not enrolled'). The non-enrollment rate like the graduation rate demonstrates a substantial relation to socioeconomic status (see column 3 of Table 2). While 22% of whites are not enrolled in spring 2001, the fraction of African Americans who are not enrolled is over fifty percent higher at 36.5%. The fraction not enrolling more than doubles across the range of household income and parental education: from less than 13% for parents with post-graduate work to more than 30% for those with no more than a high school degree and from 14% in the highest income category to 32% in the lowest income category.

Nevertheless, these data indicate that persistence at the six year mark is widespread. The first row of column 2 indicates that 13% of the entire sample is continuing to work towards a degree, meaning that 36% (13/(13+23)) of those who have not graduated are persisting. Results are similar when we define persistence as enrollment at any time in the last academic year, with persistence rising to about 40% of non-enrollment.[5] The fraction persisting is furthermore usually higher for those from less advantaged socioeconomic backgrounds as 19% of African Americans and 17% of those with the lowest household income are still enrolled. Thus, there is evidence that the lower graduation rate observed for less advantaged populations six years following matriculation may be partially explained by their higher persistence and partially offset by higher subsequent graduation rates.

These raw statistics suggest that researchers who lump all non-graduates into one category for statistical analysis may be using an oversimplified outcome measure that underestimates long-term college success. While the BPS does not follow these students beyond their sixth year, we can look at those who were persisting at the end of their fifth year and see how they progressed in the following year. Of those who were enrolled but did not graduate in the final term of their fifth year, 26% had graduated and 52% were still enrolled at the end of year six. If the progression from year five to year six is any indication of future trends, many of those classified as persisting in year six may well complete their baccalaureate degree within a year or two.

V. RESULTS

The parameter estimates for the key socioeconomic variables we obtain from simple logit models of graduation are reported in Table 3. Other parameter estimates are available upon request. A positive coefficient indicates an increased probability of graduating. The first column presents results for the model that controls only for basic demographic characteristics. The second column provides results when also controlling for academic preparation/ability, while the third column controls for the broadest array of covariates.

As the magnitude of any effect is difficult to infer from the parameter estimates in a logit model, numerical marginal effects are reported below the coefficient estimates.[6] In nonlinear specifications such as a logit, marginal effects will differ depending upon the location of the observation in the probability distribution. Marginal effects will be larger in the center of the distribution as a movement of β in either direction will capture a larger population. Thus, it is important to select a base case for analysis that holds approximately constant the baseline probabilities. As our primary interest is in identifying the relation between socioeconomic status and college outcomes, we maintain as a base case a single, white, non-Hispanic, childless, 17 year old male from New England and a residence with a sample average unemployment rate, with a college educated parent, and an annual household income greater than $75,000 – an individual from a distinctly advantaged socioeconomic background. Academic preparation and ability are assumed to be approximately modal with the highest expected level of math being trigonometry, high school GPA being between a B and an A-, and SAT test scores falling between 800 and 1100, all for respondents with a high school degree. When including the most inclusive set of covariates, the respondent is assumed to attend a public college of average selectivity that has consistently fewer than 5,000 students; to receive some grant aid; and to begin college in the fall term. The predicted probability of graduating for an individual with these characteristics ranges from 73.4% for the base model, to 75.6% for the model controlling for academic preparation/ability, to 72.8% for the most inclusive model – thus the location in the distribution, which is so important for the interpretation of logit results, is approximately constant and the marginal effects can be reasonably compared across specifications.

The basic specification (column 1) illustrates significant differences by socioeconomic status. Focusing on the marginal effects, African Americans are 13% less likely to graduate than Whites; Hispanics are 7% less likely to graduate than non-Hispanics; first generation college students are about 11 to 14% (depending on the definition) less likely to graduate than students whose most educated parent has a college degree; and those from the lower half of the income distribution are 9-11% less likely to graduate than those from the highest income quartile, holding all else equal. These differences are somewhat smaller than the raw differentials observed in Table 2 where differences between, for example, the African American and White graduation rates do not control for ethnicity, parental education, or household income, but the differences vary by population. Thus, the difference is on the order of 20-30% lower for Hispanics and first generation college students; 35% lower for African Americans; and about 60% lower for the lowest income quartile. Income in particular is a lot less important when jointly controlling for other basic demographic characteristics and conditioning on initial enrollment, even when not taking into account measures relating to ability.

The marginal impact of socioeconomic status on graduation is, however, further reduced when controlling for academic preparation/ability (column 2). The decrease is on the order of 66% for African Americans and 45% for Hispanics. The decrease is somewhat smaller for those from the bottom half of the income distribution (18-36%) and for first generation college students (16-25%). All of these changes are greater than a standard deviation in magnitude.

Only one marginal effect remains as high as ten percentage points after controlling for academic preparation/ability, whereas previously four of six were larger than ten percentage points. Overall, the impact of high school preparation/ability is both significant and substantial. The marginal impact (not reported but available upon request) of moving either from the lowest level of math (algebra/geometry) to calculus or from a combined SAT test score of less than 800 to a combined SAT test score of more than 1100 is on the order of nine to ten percentage points. The marginal impact associated with reporting a high school GPA of A versus B- or lower is even larger at 30 percentage points! Student performance in high school is a strong proxy for student success in college – much more so than socioeconomic status.

Including the commonly used, but possibly endogenous, covariates (column 3) has only a modest impact. The marginal effects for race/ethnicity rise by about one percentage point. The marginal effects for parental education rise less. The marginal effects for income rise more. That the marginal effect of being in the lowest income quartile changes the most (2.6 percentage points) is likely because this expanded specification includes controls for financial aid receipt. However, none of these differences are over one standard deviation in magnitude and so none are statistically significant. Thus, focusing on the less inclusive specification yields the same results and avoids any taint of endogeneity.

Numerical marginal effects from the multinomial logit specification are reported in Table 4 for each specification and for each outcome. The first row indicates the predicted probability given base case characteristics. Again, these probabilities need to be similar across specifications in order to allow comparison of the marginal effects across specifications. The predicted probability of graduating ranges from 73.2% to 75.7%; the predicted probability of still being enrolled ranges from 9.6% to 11%; and the predicted probability of not being enrolled ranges from 14.7% to 16.6%. These are all of relatively comparable magnitude. Not surprisingly, the predicted marginal impact of each characteristic on the probability of graduating itself, using the multinomial logit specification, is almost exactly that generated by the logit specification. Thus, we focus our discussion on the other outcomes.

The results clearly indicate that the factors distinguishing non-enrollment from graduation and those distinguishing persistence from graduation are significantly different (p-value 0.00 for all specifications). Non-enrollment and persistence are different outcomes, and policy makers should address these behaviors separately in acting to improve college outcomes.

Looking at the results from the basic specification, there are striking differences in the predicted distribution of non-graduates by socioeconomic status. Holding all else constant, the marginal effect of being Hispanic is over twice as great on persistence (5.3%) as it is on non-enrollment (1.9%). Conversely, the marginal impact of being a first generation college student on non-enrollment is distinctly larger (11-12%) than on persistence (0 to 2.6%). African Americans and those from the lowest income strata have a somewhat higher marginal probability of persisting but also a larger relative chance of not enrolling. Overall, it appears that Hispanics who have not graduated in six years may not have given up but may be on the slow road to graduation while first generation college students may be gone for good.

The results for Hispanics and first generation college students are robust across specifications, albeit with somewhat smaller and less significant marginal effects as more controls are added. As was the case with the simple logit, the marginal effect of belonging to the lowest income quartile on graduating is much smaller when controlling for academic background. This decrease stems largely from a reduction in the marginal impact on non-enrollment. The marginal effect of coming from a low income household on the probability of persisting diminishes only slightly. Controlling for the largest set of covariates, including first year financial aid type, again increases the marginal effect of income on graduation and non-enrollment. To see if this effect could be driven by differential first year financial aid by income, interactions between income and aid type were incorporated in the specification. These terms were neither jointly nor individually significant. The marginal impact of being African American on outcome probabilities declines precipitously after controlling for academic ability. Even though the marginal effect of being African American on non-enrollment remains greater than the marginal effect on persistence, neither impact is statistically significant at conventional levels.

Educational background continues to have the same large marginal impact on the probability of graduating that it had using the simple binary outcome measure. Thus, the marginal impact associated with the rigor of the student's curriculum and with the student's test score is on the order of nine to ten percentage points, while the marginal impact associated with a change from the lowest to the highest high school GPA is around 30 percentage points. Of greater interest is how the marginal impact of academic background differs for those who have not graduated. While each of our measures appears to have a statistically significant marginal effect on both persistence and non-enrollment, there are some differences. The predicted marginal effect of high school GPA is about three times as large upon non-enrollment as upon persistence. Jointly the coefficients to our high school GPA measures are significantly different for non-enrollment than for persistence at the one percent level. By contrast, test score measures appear to have a greater marginal impact on persistence than on non-enrollment. While those with both higher and lower than median test scores have significantly different probabilities of persisting, only those with lower test scores are significantly more likely not to be enrolled.

Finally, while the rigor of the high school curriculum has three times the predicted marginal impact on non-enrollment as on persistence, these effects are not significantly different from zero. Thus, the multinomial logit specification highlights some differences in how academic background is associated with six-year persistence and dropout behaviors. These differences further emphasize the importance of utilizing a multinominal logit versus the traditional binominal logit to evaluate college outcomes.

To test the robustness of our results and to see if any patterns arise using different observation windows, we reran the analysis (1) coding respondents enrolled at any point during the sixth year as persisters and (2) using fifth year (Spring 2000) outcomes (results available upon request). Obviously, a smaller fraction has graduated in five years as compared to six (58% versus 63%). While 20% were still enrolled in spring 2000 (year five), 16% were enrolled at some point during the 2000-2001 academic year, and 13% were still enrolled in spring 2001. The fraction classified as having withdrawn is relatively stable, ranging from 22% in year 5 to 23% in year six. This stability arises because most of those classified as withdrawals have not been enrolled for three years and 40% have not been enrolled for four years. The majority are long term dropouts. Reestimating the multinomial logit model with these alternative definitions of the dependent variable does not substantially change our results. If anything they show that academic background explains a greater share of the graduation rate differential at the five than at the six year cutoff. This result may be due to the fact that as students persist, their high school record matters less.

We also tested for interaction effects between race/ethnicity and income/parental background. No significant interaction was identified. From this we can infer that the effects of low income or first generation status are not different by race or ethnicity.

VI. CONCLUSION & DISCUSSION

Lower socioeconomic status has long been associated with a failure to complete college. In this study we make two primary contributions. First, we examine the relation between commonly used socioeconomic factors and graduation. In doing so, we are able to include a broader array of controls for individual academic preparation/ability than is typically possible. This approach allows us to assess the impact of socioeconomic status on college outcomes, holding academic background constant. Second, and significantly different from prior studies, we distinguish between those non-graduates who are still enrolled six years following matriculation and those who are not still enrolled. Standard logit analysis treats all non-graduates the same, and hence in some sense as failures. We find that 36% of those who had not graduated in six years were still enrolled when last observed. Persistence at the six year point is substantial. Furthermore, our results indicate that persistence and non-persistence are statistically distinct outcomes. Evidence from those persisting at the five year mark suggests a good fraction of those still enrolled after six years may in fact go on to graduate. Thus, persisters may just be taking longer to graduate. If students from more disadvantaged backgrounds are disproportionately likely to persist, significant differences in graduation rates by socioeconomic status may disappear over time.

Using a national sample of first time undergraduates matriculating in 1995/96, we find that simply jointly controlling for basic demographic (primarily socioeconomic) characteristics explains a substantial fraction of the raw graduation rate differences reported by socioeconomic status. There is a lot of overlap in terms of income, parental education, race, and ethnicity. Still, the differences in graduation rates remain substantial, typically over ten percentage points.

Adding controls for academic background as measured by test scores, high school grades, and high school curriculum reduces those adjusted graduation rate differences by about two-thirds for African Americans, half for Hispanics, and 20-35% for first generation college students and students from low income households. Thus, the predicted difference by race/ethnicity falls to about four percent and becomes at best marginally statistically significant. Those from the lowest half of the income distribution are predicted to be seven percent less likely to graduate relative to those from the upper income quartile. First generation college students are predicted to be between nine and 11% less likely to graduate than students with more highly educated parents. These income and first generation six-year graduation differentials are substantial and statistically significant. While academic background is substantially and significantly associated with college graduation, controlling for academic background does not eliminate all observed differences in graduation rates by socioeconomic status.

We then extend the standard analysis of college outcomes in a novel way to distinguish among three outcomes: graduation, continued enrollment, and non-enrollment. Using a multinomial logit specification, we find evidence that treating all those who have not graduated as a simple, single population is not statistically appropriate. This more complex analysis reveals significant differences in the marginal impact of socioeconomic status on the probability of persisting. Those of Hispanic descent are significantly more likely to persist than non-Hispanics but are not significantly more likely to stop enrolling. Conversely, first generation college students are significantly more likely to not be enrolled, but not significantly more likely to persist than non-first generation college students. African American students and those from lower income households have higher probabilities of both persisting and not enrolling, as compared to their white and higher income counterparts. Although academic background is an important predictor of college outcomes, controlling for academic background and other covariates does not substantially change this persistence story.

Equal access to higher education has been a social goal for decades now in the United States. Attention has more recently shifted from access to persistence and degree receipt. These outcomes are important for institutions, educators, and policy makers both because limited resources make time spent in school expensive and because it is success in college, not just access, that will help us achieve greater social equality. Most research on persistence has focused on the early years of the college experience, commonly the first to second year transition. Research on degree receipt has focused on six-year graduation rates. That focus on degree receipt fails to distinguish between persisters and non-persisters at the six-year mark. Our analysis begins to fill that substantial void and suggests that long term persistence is deserving of further attention. The fact that many students who are persisting at the five year mark successfully complete their degree in six years is promising, but data that follow students beyond the six year window are needed to determine if those persisting at the six year point actually do graduate. The higher average persistence rate of the Hispanic population also requires some analysis, particularly if policy makers wish to speed time-to-degree for such students.

A common thread throughout this discussion is the importance of academic preparation in the K-12 years on college success. Colleges work with the raw material they receive, and it is costly for colleges to change those K-12 preparations. From a broad policy perspective, improvements in K-12 education are important not only as the United States needs to be more competitive in the global economy, but also because such improvements should improve college success. Success is important especially for those subgroups in society that have been historically underrepresented in college and hence have only slowly advanced up the income and education social ladder. Policy makers should consider the impact on college graduation and persistence as they evaluate the benefits of improving K-12 education in general. We do, however, continue to observe significant differences in college outcomes by income and family background even after controlling for academic background. This suggests that changes at the college level to help those historically underrepresented may also be in order. Socioeconomic status, due to the luck of the draw at birth, remains a barrier to college completion that needs to be eliminated, if one believes in equal opportunity for all.

Table 1
Sample Means
(% except where noted)

	Mean	Std. Dev.
Basic Specification		
Female	0.550	0.498
White	0.776	0.417
African American	0.109	0.311
Other race	0.115	0.320
Hispanic	0.083	0.276
Parental Education		
High school	0.305	0.012
Some college	0.124	0.329
College	0.251	0.434
Post-graduate	0.264	0.441
Missing	0.055	0.229
Family Income		
Independent	0.028	0.166
Income ($000s)	60.648	54.651
< $25,000	0.224	0.417
$25-$50,000	0.262	0.440
$50-$75,000	0.245	0.430
>= $75,000	0.269	0.443
Age - 17	1.412	0.756
Ever married male	0.004	0.063
Ever married female	0.007	0.083
Father	0.004	0.061
Mother	0.010	0.101
Unemployment rate in state of residence	5.494	1.194
Measures of Academic Preparation/Ability		
No high school diploma	0.011	0.103
Highest level of math:		
Algebra II or less	0.229	0.420
Trigonometry	0.163	0.370
Pre-calculus	0.230	0.421
Calculus	0.259	0.438
Missing	0.119	0.324
Standardized Test Information		

SAT score of 800-	0.186	0.389
SAT score of 800-1000	0.468	0.499
SAT score of 1100+	0.317	0.465
Took ACT test	0.306	0.461
Missing test score	0.029	0.169
High school GPA		
B- or lower	0.088	0.283
B- to B	0.142	0.349
B to A-	0.270	0.444
A- or higher	0.384	0.486
Missing	0.117	0.322
<u>Other Covariates</u>		
Public institution	0.642	0.479
Barron's Admissions Competitiveness Index 1992		
Less selective	0.259	0.438
Moderately selective	0.412	
Very selective	0.328	0.470
Growth in FTE undergraduates (1992-1996 average)		
Negative growth (-1%-/year)	0.310	0.462
No growth	0.410	0.492
Positive growth (1%+/year	0.280	0.449
Institution size		
Number of undergraduates	10398	8630
< 5,000	0.346	0.476
5-10,000	0.237	0.425
10-20,000	0.278	0.448
> 20,000	0.139	0.346
Began in the Spring not Fall term	0.043	0.005
Financial Aid		
Received a loan	0.497	0.500
Received a grant	0.621	0.485
Received work study	0.166	0.372
Number of Observations	~5820	

Eight regional dummies are also incorporated in each specification.

Table 2
Raw Outcomes by Socio-Economic Status

	Six Year Outcome Probabilities		
Sample	Graduate	Still Enrolled	Not Enrolled
Full	63.23	13.36	23.41
Race			
White	65.60	12.33	22.07
African American	44.65	18.80	36.55
Other	64.85	15.11	20.04
Ethnicity			
Non-Hispanic	64.08	12.75	23.17
Hispanic	53.91	20.02	26.07
Parental Education			
≤ High School	50.07	16.58	33.36
Some college	55.53	12.99	31.48
College	69.27	12.51	18.22
Post-graduate	76.97	10.39	12.64
Income			
< $25,000	50.82	17.44	31.73
$25-$50,000	57.52	14.00	28.47
$50-$75,000	66.88	12.76	20.36
≥ $75,000	75.81	9.87	14.32
Number of Observations	~5820		

Table 3
Impact of Socioeconomic Status on Six Year Graduation Rate
Results from a Logit Model

	Base Case		With Academic Preparation/Ability		Largest Set of Covariates	
	Coefficient		Coefficient		Coefficient	
African American	-0.5955	***	-0.2172		-0.2622	*
	(0.1324)		(0.1365)		(0.1452)	
	-13.09%		*-4.22%*		*-5.49%*	
Hispanic	-0.3440	**	-0.2039		-0.2217	
	(0.1385)		(0.1507)		(0.1611)	
	-7.25%		*-3.95%*		*-4.60%*	
Parental Education						
≤ High School	-0.6543	***	-0.5257	***	-0.5071	***
	(0.0783)		(0.0796)		(0.0819)	
	-14.51%		*-10.92%*		*-11.09%*	
Some College	-0.4924	***	-0.4385	***	-0.4357	***
	(0.1350)		(0.1331)		(0.1340)	
	-10.65%		*-8.95%*		*-9.42%*	
Post Graduate	0.2846	**	0.1837		0.1523	
	(0.1192)		(0.1298)		(0.1293)	
	5.20%		*3.23%*		*2.91%*	
Household Income						
< $25,000	-0.4281	***	-0.2975	**	-0.3966	***
	(0.1435)		(0.1395)		(0.1535)	
	-9.16%		*-5.89%*		*-8.51%*	
$25-50,000	-0.4952	***	-0.4305	***	-0.4918	***
	(0.1183)		(0.1188)		(0.1260)	
	-10.72%		*-8.77%*		*-10.73%*	
$50-75,000	-0.2595	*	-0.1790		-0.1806	
	(0.1334)		(0.1350)		(0.1450)	
	-5.38%		*-3.45%*		*-3.72%*	

Standard Errors in parentheses. Marginal effect reported below.

Asterisks indicate significance level: *** 1%, ** 5%, * 10% for a 2-tailed test.

All specifications include controls for gender, other race, independence from parents, age-17, region of residence, the unemployment rate in the state of residence, and gender-specific marital and parental status.

Academic preparation/ability measures include controls for highest math expected in high school, high school GPA, SAT equivalent test scores, and high school degree receipt.

The largest set of covariates includes the type of first year financial aid received; a dummy to identify those who first enter in the spring term; college type (public/private), selectivity, growth rate, and size.

Table 4

Marginal Impact of Socioeconomic Status on Three Six Year Outcomes

Results from a Multinomial Logit Model

	Base Case			With Academic Preparation/Ability			Full Set of Covariates		
	Graduated	Still Enrolled	Not Enrolled	Graduated	Still Enrolled	Not Enrolled	Graduated	Still Enrolled	Not Enrolled
Base Probability	73.35%	10.06%	16.59%	75.72%	9.57%	14.71%	73.17%	10.99%	15.85%
African American	-13.05%	4.30%	8.75%	-4.21%	1.55%	2.66%	-5.55%	2.65%	2.90%
	(0.0000)	(0.0540)	(0.0010)	(0.1350)	(0.3880)	(0.1980)	(0.0960)	(0.3020)	(0.1890)
Hispanic	-7.24%	5.31%	1.93%	-3.84%	3.55%	0.28%	-4.54%	4.08%	0.46%
	(0.0250)	(0.0050)	(0.4250)	(0.1980)	(0.0420)	(0.8970)	(0.1890)	(0.0640)	(0.8470)
Parental Education									
<= High School	-14.71%	2.65%	12.06%	-11.10%	2.05%	9.06%	-11.11%	2.08%	9.04%
	(0.0000)	(0.0450)	(0.0000)	(0.0000)	(0.1090)	(0.0000)	(0.0000)	(0.1780)	(0.0000)
Some College	-10.94%	0.13%	10.81%	-9.31%	-0.05%	9.36%	-9.50%	-0.19%	9.69%
	(0.0010)	(0.9310)	(0.0000)	(0.0040)	(0.9740)	(0.0020)	(0.0070)	(0.9050)	(0.0030)
Post Graduate	5.31%	-1.19%	-4.12%	3.32%	-0.64%	-2.68%	2.95%	-0.42%	-2.52%
	(0.0180)	(0.3010)	(0.0280)	(0.1460)	(0.5670)	(0.1290)	(0.2200)	(0.7470)	(0.1480)
Household Income									
< $25,000	-9.23%	3.57%	5.65%	-5.95%	3.07%	2.88%	-8.64%	3.72%	4.92%
	(0.0020)	(0.0590)	(0.0190)	(0.0290)	(0.0960)	(0.1840)	(0.0040)	(0.1100)	(0.0540)
$25-50,000	-10.75%	2.31%	8.44%	-8.83%	2.01%	6.82%	-10.76%	2.51%	8.25%
	(0.0000)	(0.1530)	(0.0000)	(0.0000)	(0.1860)	(0.0000)	(0.0000)	(0.1760)	(0.0000)
$50-75,000	-5.37%	2.01%	3.36%	-3.49%	1.48%	2.02%	-3.80%	1.48%	2.32%
	(0.0500)	(0.1150)	(0.1830)	(0.1760)	(0.2250)	(0.3690)	(0.1850)	(0.2970)	(0.3480)

P-values in parentheses. The models correspond to those estimated for the logit specification.

The base probability is for a single, childless, 17 year old white, non-Hispanic, non-first generation male with a household income of > $75,000, who lives in New England, in a state with a sample average unemployment rate.

The base probability for academic preparedness and ability is for an individual who has a high school diploma, expects to complete trigonometry, has an A average in high school, and has an SAT score of 800-1100,

The base probability for the full model is for an individual who receives no financial aid, enters a moderately selective public institution with a constant size of less than 5000 students in the fall term

References

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

Altonji, J. G. "The demand for and return to education when education outcomes are uncertain." *Journal of Labor Economics* 11 (1), 1993, 48-83.

Barrow, L. and C. E. Rouse. "Do Returns to Schooling Differ by Race and Ethnicity?" *American Economic Review*, 95(2), Papers and Proceedings, 2005, 83-7.

Bean, J. P. "Dropouts and turnover. The synthesis and test of a causal model of student attrition." *Research in Higher Education*, 12(2), 1980, 155-87.

Becker, G. *Human capital*. New York: National Bureau of Economic Research, 1964.

Belley, P. and L. Lochner. "The Changing Role of Family Income and Ability in Determining Educational Achievement." National Bureau of Economic Research Working Paper 13527, 2007.

Bound, J., M. F. Lovenheim, and S. Turner. "Why have college completion rates declined? An analysis of changing student preparation and collegiate resources." *American Economic Journal: Applied Economics*, 2(3), 2010, 129-57.

Carneiro, P. and J. J. Heckman. "The Evidence on Credit Constraints in Post-Secondary Schooling." *The Economic Journal*, 112(482), 2002, 705-34.

Cameron, S. V., and J. J. Heckman. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, 109(3), 2001, 455-99.

The College Board. "Concordance Between SAT I and ACT Scores for Individual Students." Research Notes. Office of Research and Development. RN-07, June 1999.

The College Board. The College Completion Agenda: 2010 Progress Report. 2010.

Cragg, K.M. "Influencing the probability for graduation at four-year institutions: A multi-model analysis." *Research in Higher Education*, 50(4), 2009, 394-413.

Deming, D and S. Dynarski. "Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor." National Bureau of Economic Research Working Paper 15387, 2009.

DesJardin, S.L., D.A. Ahlburg, and B.P. McCall. "An Event History Model of Student Departure." *Economics of Education Review*, 18(3), 1999, 375-90.

Dynarski, S. M. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *The American Economic Review*, 93(1), 2003, 279-88.

Hu, S., and E. P. St. John. "Student persistence in a public higher education system: Understanding racial and ethnic differences." *Journal of Higher Education*, 72(3), 2001, 265-86.

Johnson, I. "Enrollment, persistence and graduation of in-state students at a public research university: Does high school matter?" *Research in Higher Education*, 49(8), 2008, 776-93.

Kane, T. J. "College entry by blacks since 1970: The role of college costs, family background, and the returns to education." *Journal of Political Economy*, 102(5), 1994, 878-911.

Kuh, G. D., J. Kinzie, J. A. Buckley, B. K. Bridges, and J. C. Hayek. *What matters to student success: A review of the literature*. Final report for the National Postsecondary Education Cooperative and National Center for Education Statistics. Bloomington, IN: Indiana University Center for Postsecondary Research. http://nces.ed.gov/npec/papers.asp. 2006.

Manski, C. F. "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon." *Economics of Education Review*, 8(4), 1989, 305-12.

Schmitt, C.M. "Documentation for the Restricted-Use NCES-Barron's Admissions Competitiveness Index Data Files: 1972, 1982, 1992, 2004, and 2008." National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C., NCES 2010-330, 2009.

Stratton, L. S., D. M. O'Toole, and J. N. Wetzel. "A multinomial logit model of college attrition that distinguishes between stopout and dropout behavior." *Economics of Education Review,* 27(3), 2008, 319-31.

Vignoles, A. F. and N. Powdthavee. "The Socioeconomic Gap in University Dropouts." *The B.E. Journal of Economic Analysis & Policy*, 9(1), Article 19, 2009.

Footnotes:

* This material is based upon work supported by the Association for Institutional Research, the National Center for Education Statistics, the National Science Foundation, and the National Postsecondary Education Cooperative, under Association for Institutional Research 2010 Grant Number RG10-128. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Association for Institutional Research, the National Center for Education Statistics, the National Science Foundation, or the National Postsecondary Education Cooperative. We thank participants at the 2011 AIR Forum and the University of Cergy Pontoise for helpful suggestions.

Leslie S. Stratton, Professor, Department of Economics, Virginia Commonwealth University, 301 West Main Street, PO Box 844000, Richmond, VA 23284-4000. Phone 804-828-7141, Fax 804-828-9103, E-mail lsstratt@vcu.edu. Research Fellow at IZA, Bonn, Germany.

James N. Wetzel, Professor, Department of Economics, Virginia Commonwealth University, 301 West Main Street, P.O. Box 844000, Richmond, VA 23284-4000. Phone 804-828-7145, Fax 804-828-9103, E-mail jwetzel@vcu.edu.

[1] Bound, Lovenheim, and Turner (2010) report evidence from other sources that completion rates have fallen.

[2] See Altonji (1993) for a model of such decision making under uncertainty and Manski (1989).

[3] See Kuh et al. 2006 for a review

[4] NCES Security restrictions require we round sample sizes to the nearest ten.

[5] To assess the degree to which our results might be sensitive to our definition of persistence, we looked more closely at enrollment records. We find that about 50% of those we classify as not enrolled have enrolled for no more than two years of study in the six years they are observed. They either dropped out, never to return, or floated in and out of college. By comparison, only 3% of those classified here as persisters have completed as few as two years of study. On average the enrollment patterns of these individuals are quite different. Nevertheless, we report below estimates using alternative definitions to test the sensitivity of our results to our chosen definition of persistence and to our chosen window of analysis (six years following matriculation).

[6] Analytic marginal effects are similar and available upon request.

The “Art and Science” of Developing Institutional Peers: Methods of Exploring and Determining Peer Institutions

Joanna Musial-Demurat
Columbia Green Community College
State University of New York

Bruce Szelest
University at Albany
State University of New York

Abstract: Dimensions of finance, size, complexity, and quality are captured in the peer group development process used by a public research university. In addition to contrasting traditional rank/distance methods with a cluster analysis approach, data integrity/limitations and campus political considerations shape the “art and science” of developing institutional peers.

Paper presented at the 2011 North East Association for Institutional Research, Boston, Massachusetts

Introduction

The State University of New York (SUNY) at Albany is one of 64 campuses in the state university system. Although there is a high diversification of institutions in the SUNY system, it nonetheless does not provide a natural or traditional peer group at the research university level, as there are only four university centers within the SUNY system. We therefore needed to develop a list of peer institutions that could be used in for our institutional planning, as well as within system-wide planning processes. A list of Albany peers was successfully developed by the campus in 1996, and then later in 2006, and peer benchmarking has been used continuously for various purposes. In this age of demand for accountability, transparency, and accreditation, as well as intensified competition and the pursuit of multiple initiatives, colleges and universities are constantly adjusting and transforming themselves to meet new challenges and enhance their standing. These changes sometimes influence important characterizes of higher education institutions, and thereby the selection of appropriate benchmarking peer groups. Thus we have decided to reevaluate and retest the peer groups that were created via cluster and rank distant analyses in 2006.

Literature Review

Only some higher education institutions have a natural or traditional peer group, like the Ivy League Colleges, those in the Big Ten, or the HEDS Consortium. Most institutions do not have such a convenient comparison group and have to develop a set of peers for benchmarking their performance, or the pursuit of desirable traits. During this process an important question needs to be answered: how to identify and evaluate candidates for inclusion in the peer group? There is no one generally accepted standard technique with which to identify peer institutions. Teeter and Brinkman (1987) suggest that both technical and political considerations contribute to the development of peer comparisons.

The authors (1992) emphasize that before selecting comparison groups, it is crucial to understand political aspects of using comparative data. The issues to be addressed help to determine the type of comparison group to be chosen. The authors describe four major types of comparison groups; 1) competitors, which regard to applicants, faculty, or financial resources, 2) aspirational, those institutions we strive to be like in some respects, 3) predetermined, those institutions that are natural, traditional, or which share a common jurisdictional area, and 4) peers, which can be used in benchmarking. This classification of comparison groups is very useful for institutional researchers in identify the pool of institutions that should be selected based on the given situation and purpose of the comparison.

Methods of selecting peer institutions vary with the emphasis given to data, statistics, and input from political players. Brinkman and Teeter (1987) describe four peer selection methods starting with a cluster analysis approach which is based on only data and statistics, through a hybrid approach which combines data, statistics, and judgment, followed by a threshold approach that has a combination of data and judgment, and panel review based only on judgment. These methods range from a purely statistical technique on one end (like cluster analysis) to mainly subjective methods on the other end (like panel reviews).

All these methods are used by institutional researchers, but in recent years, several scholars have strongly recommended using a hybrid approach. The main advantage of this approach is that it incorporates both data driven analysis and expert judgment (Ingram 1995; Zhao 1997; Lang 2000; Xu 2008; Archer, Reiss, Armacost, Sun, Fu 2009). As defined by Teeter and Brinkman (1992) the hybrid approach combines a strong emphasis on data and on input from administrators, with statistical algorithms for analyzing data. It is one of the most popular methodologies in peer selection (Teeter and Brinkman 1992), and is discussed in detail in the present paper.

The organizational literature in general, and studies of higher education in particular, indicate that there are important differences within and between organizations based upon their mission, size, wealth, complexity, and quality. Thus these dimensions-- especially size, finance, quality, and complexity-- have been commonly used for peer analysis purposes in studies conducted in the last fifteen years (Szelest, 1996; Zhao 1997; Weeks, Puckett, Daron 2000; Xu 2008; Gaylor 2009; Nzeukou and Muntal 2010). Other dimensions include reputation, growth and mission, geographic location, research activity, and price.

After making decisions on the type of comparison group, methods of selecting institutions, and dimensions for peer analyses it is time for another very important step: the choice of variables. According to Aldenderfer and Blashfield (1984), the selection of variables in peer analyses is one of the most critical steps in the research process. The choices of characteristics used for peer comparisons strongly influences the specifications of institutions and thereby the selection of a peer group. A broad range of variables starting from basic ones like enrollment or, student mix, through financial variables, admissions selectivity, etc., represent institutional dimensions which are commonly used in conducting comparative analyses of peers. Peer groups are often formed based upon combinations of these variables. Some researchers like Nzeukou and Muntal (2010) use a large number of variables grouped in a few dimensions in their analysis aiming to select peer institutions. Other researchers like Archer, Reiss, Armacost, Sun, Fu (2009) or Weeks, Puckett, Daron (2000) focus on fewer variables while selecting peer groups. There is no consensus among scholars on the right type and the correct number of variables which should be used for peer comparison analyzes.

Not only is the identification of selection variables influenced by subjective judgments but also the possible assignment of variables weights (Zhao, 1997; Xu, 2008). Some scholars (Szelest, 1996; Xu, 2008) have made conscious decisions not to weight the variables/factors in their studies. One given reason rests on the assumption that variables/ factors may receive special considerations in pre-final analysis procedures so they do not need weighting in the final stage of peer comparison. For example, size varable is frequently considered in the initial processes of peer comparisons and is captured through multiple measures, so it may not need additional weight in the final step. Other researchers, such as Week et al (2000), have made a decision to weight the selection variables to give greater or lesser emphasis to key factors related to each campus's mission and programs. Lang emphasizes that in some instance good peer choices may be eliminated from a peer list because a heavily weighted variable may cause the institution to appear too different from the target university. Thus the weighting needs to be used with due reflection.

This present research surveys an additional fifteen years (1996-2011) of scholarly literature. It uses best practices around the most frequently used methodological approaches, dimensions upon which comparisons are often based, and the individual variables and metrics used in identifying peers. Overall, the structure of this research is based on methods and findings presented in the paper "In Search of Peer Institutions: Two Methods of Exploring and Determining Peer Institutions" (Szelest, 1996). That paper is a core starting point for this research in terms of the preliminary selection of variables and methodological approach. Our research mimics two aspects of the previous selection of variables used for the final analyses in the article from 1996. First, individual measures selected for inclusion were made to anticipate campus decision makers' interests and to reflect their concerns. Secondly, the selection of variables is based on their appropriateness in capturing dimensions of finance, size, complexity, and quality. In terms of methods of analyzing variables, the research uses a combination of statistically driven techniques, a number of subjective judgment methods in regard to variable and measure selection, and weighting schemes.

Dataset Development

A dataset was developed that included 134 public institutions (4-year or above, offering the doctorate) for potential analysis. A decision was made early in the process to consider only those institutions with a strong emphasis on doctoral education. Therefore, those institutions that do not offer the doctorate are excluded from analysis. Because the target institution to which the peer selections are to be made is the University at Albany, a Carnegie Research University with *very high research activity* (2000), potential peer institutions with this same classification were chosen for exploration. In addition, because prior peer selection experiences in 1989, 1996, and in 2006 have included Research Universities subsequently classified with *high research activity* by the Carnegie Foundation, these institutions were desirable to include in the initial list of potential institutions. Extraction of institutions with these criteria restrictions (public, 4-year or above, offering the doctoral degree, and with *very high* or *high* research intensity) from the IPEDS data analysis tool resulted in an initial dataset of 141 institutions. The comparison group is restricted to public institutions due to the public control of the target institution, which is part of the State University of New York (SUNY) system.

Additionally, potential institutions include medical and engineering schools although the University at Albany does not offer such programs. Owing to the fact that previous peer analyses have included several institutions which offer medical or engineering degrees, and the fact that the vast majority of public doctoral research universities offer engineering programs, it was determined early on that these attributes in and of themselves should not disqualify institutions from the analysis. That said, the presence of engineering or medical schools can indeed exert effects on the variables upon which the peer selections are made, particularly those in the financial realm like faculty salaries. Therefore, while we do not exclude institutions from analysis on this basis, we are sensitive to this institutional attribute and allow for noting the presence or absence of engineering and medical programs in the final peer list. A similar situation exists with regard to whether or not an institution has land grant status. We did not make this a requisite criterion either, as the University at Albany is not a land grant institution, but several land grant institutions in the past have been considered among Albany's peers. The key to moving past this consideration and proceeding with the analysis is to acknowledge that driving similarities may lie in other critical dimensions, thus over-riding these particular institutional attributes.

Missing data are a concern in any quantitative analysis. This is true even with IPEDS data for which NCES has the power to impose monetary (and worse!) penalties for institutions that do not submit timely or accurate data. There are no doubt myriad reasons why data may show up as missing in IPEDS, including applicability, reporting errors, or even a result of extracting data from one particular IPEDS survey like the GASB financials survey as it appears that public universities should use that specific reporting form, but in reality a handful of public institutions use the FASB financial instrument. As discussed below, the analytic methods used in this research require complete data. Missing data excludes institutions from the analyses. Therefore, in instances when missing data exist, the researchers explored individual cases and sought out proxy data from other sources. These instances were few, but noteworthy. For example, it appears that Penn State University, the University of Pittsburgh, Temple University, and the University of Delaware used the FASB financial form, which is primarily for private not-for-profit institutions, than the GASB form is used by just about every other public research university. In addition, while 2009-10 data were available in the GASB data, the most recent FASB data available were 2008-09 data. Rather than exclude these potential peers from the analysis, or resort to mean substitution of data, we used the 2008-09 financial figures for these institutions but inflated them by 2.3 percent, in step with the 2008 to 2009 higher education price index (HEPI).

Mean substitution was used as a last resort when missing data were found in the metrics used to in the analyses. In these cases, only a handful of institutions were affected. The rationale for using mean substitution is that because of the number of measures used (78), it is believed that using the mean value on one or two will not adversely affect the analyses. We included a flag in the dataset to mark those measures (and institutions) in which mean substitution was used so that we could revisit the details of the data should these institutions find their way into the final peer listing. Finally with respect to missing data, it is worth noting that in other instances when missing data were observed, we made the decision to simply remove the institution from the analysis if it had inherent characteristics that resulted in a very low probability for inclusion, given the various attributes of the target institution. An example of these institutions would be the CUNY Graduate School, which has an overall enrollment of about 6,000 students, with only about 1,000 of them being undergraduate – as compared to Albany's 18,000 students, and about 12,000 of whom are undergraduates. Another notable institution dropped from the analysis due to missing data was Rutgers-Newark. This campus had quite a few missing measures, given the data elements that were extracted from IPEDS for this study. Rather than invest considerable time and energy into finding the actual data or developing proxy data, examination of the Rutgers-Newark profile showed that compared to our target institution, it is only about half of Albany's size, thus making it highly unlikely that campus academic leaders would have considered it a viable peer institution in the first place. With the dataset in place, we then moved into the analysis realm.

Variables Development

After selecting only public doctoral research universities (with *very high* or *high* research activity), we chose variables based on their appropriateness in capturing dimensions of finance, size, complexity, and quality. We based our selection of variables on the reevaluation of variables used in Szelest's study (1996) and examination of the most popular variables used by other researchers in the last fifteen years. The final analysis includes seventeen financial, twelve size, five quality, and twenty-five complexity measures. The complexity measures overwhelming focus on undergraduate and graduate programmatic mix.

Finance

The measures of institutional finance address both overall support and more specific expenditure functionality. We compute total expenditures by adding total operating and non-operating expenses. We include non-operating expenses like investment income, etc. as these seem to be in practice used to support the institution. Some institutions like the Pennsylvania State University, Temple University, and the University of Delaware use a different IPEDS form (FASB rather than GASB) that utilizes this more encompassing figure. Similarly, for calculation of total revenues we include total revenues from operating and non-operating revenues. In terms of percent of budget from state support we combine state appropriations and tuition and fee income as "state support" because they are often intermingled, especially in NYS, in terms of how public higher education institutions are funded. Furthermore, we calculate state support per student by adding total operating and non-operating revenues divided by total student FTE. This variable measures the state resources directed in support of each student within overall institutional funding; state aid to students (e.g., TAP in NYS) is captured below in the variable for state commitment to providing student aid. Aggregate expenditures on organized research are used to capture magnitude of the research operation. Evaluation of the relative degree of research emphasis in the operating budget provides us with context of the relative nature of the research mission within the context of the entire university. Specific expenditure account categories of instruction, public service, academic support, student services, and institutional support are evaluated in terms of their share of budget. Thus each specific expenditure is divided by the total operating and non-operating revenues. IPEDS sourced average salary of faculty members with various ranks equated to a 9/10 month appointment is used as a proxy for faculty financial support. Finally, state and institutional commitments to providing student aid, also obtained from IPEDS, help us measure the financial support received by students at potential peer institutions.

Size

The size variables are categorized in three groups: variables which measure the size of student body, size of the faculty body, and the number of degrees awarded. Total student FTE and headcount enrollment, as well as graduate headcount, are used to capture both size and mission. Headcount enrollment is reported by IPEDS and FTE students are calculated by combining full-time headcount and one-third of the part-time headcount enrolment. Additionally, the size of entering freshmen and transfer cohorts are included in the analyses. In terms of faculty body we evaluate the number of faculty members and non-faculty employees, which allows comparisons of instruction/research/public service FTE staff and thereby illustrates some aspects of college mission. The total number of all degrees awarded, as well as the number of doctoral degrees awarded, help measure one aspect of institutional output, graduates. All size data come from IPEDS.

Complexity

The twenty-five complexity measures address administrative complexity, student body composition, graduate emphasis, and the residential nature of the campus. We use three individual measurements of administrative complexity. Full-time to part-time faculty mix, faculty to non-faculty employee mix, and per faculty production of doctoral graduates help gauge overall academic complexity. Commonly accepted indicators of student body complexity include socio-economic distribution measured based on the percent of freshmen Pell recipients, age distribution as a proportion of undergraduates of traditional age (i.e., 18 to 24 year olds), and diversification of the student body defined by the percentages of minority and international students. Emphasis on graduate education derives from the percent of the student body that is graduate. Lastly, the residential nature is measured by the percentage of students living in campus dormitories determined by dorm capacity divided by student body.

Quality

Unfortunately, it is not easy to directly verify the quality of higher education institutions due to a lack of well-defined and measureable indicators. Therefore, for the quality dimension we do not have ”true” quality indicators, but then again, no one else does either! Rather, we focus on evaluation of quality input and output variables. The input measurements examine two main concepts: *academic preparedness of the student body* calculated based on SAT scores (and ACT scores converted to their SAT equivalent) and *institutional selectivity & attractiveness of students* calculated based on freshman admit and yield rate. In terms of output variables we include two measurements: full-time freshmen student retention rate and 6-year graduation rate for the most recent calculable freshmen cohort. Additionally, we measure quality of education also by student faculty ratio, which is computed by dividing IPEDS derived student FTE by the number of full-time faculty reported in IPEDS.

Methodology

As in previous peer institution explorations, we conducted a number of preliminary examinations by looking at institutional attributes like Carnegie status, the presence or absence of medical or engineering schools, as well as by various demographic and ecological characteristics. In the past, we have found that surveying the institutional landscape and mapping out institutional typologies to be an excellent means of educating campus stakeholders about institutional similarities and dissimilarities.

For example, two important variables to deans and the campus executive officers are the campus's research profile, and state support. The latter factor is especially of note, as Albany has experienced cumulative state appropriation reductions totally about 30 percent between 2008 and 2011. Therefore, campus leaders are eager to communicate to external constituents, like state legislators and planning officers in the executive branch, just how successful the University is in securing external funds in carrying out its research mission. The typology chart on the next page categorizes institutions by their percentage of operating budget (tuition and fees plus state appropriation) and by the research and development expenditures they generate in science and engineering. For illustration purposes, we arrange institutions by quintile on each measure. While quartiles or other percentile groupings are possible, the decision was made to use quintiles as they appear to be a convenient compromise by showing a bit more differentiation than quartiles, yet have considerably less precision (and require considerably less mental processing) than deciles.

Quintile Groupings of State Support to Operating Budget versus Science & Engineering R&D Expenditures

Magnitude of Science & Engineering R&D Expenditures	1	2	3	4	5
5	GA T OH SU U CA Davis U CA LA U CA San D U FL U IL Urbana U KY U MI Ann Arbor U NC Chapel Hill U TX Austin U WA Seattle U WI Madison	PA SU TX A&M U CA Berk U MN Twin Cit U NE Lincoln	LA SU A&M MI S U Purdue U U AZ U Cinci U MD College Pk VA Polytech	NC SU Raleigh U Pittsburgh	
4	CO S U **Stony Brook** U CA Irvine U CO Denver U IA U IL Chicago U MO Columbia U UT U VA	IUPUI MS SU OR SU U CO Boulder U CT U HI Manoa U So FL U TN **UALBANY**	IA SU U KS WA SU	Rutgers New Br U CA Santa B U GA Wayne SU	AZ SU **U Buffalo**
3	U AL Birmingham U NM	MT SU NM SU U CA Santa Cr U Louisville U NH U NV Reno U OK Nornam UT SU	Clemson U Fl SU KS SU OK SU U AK Fairbanks U AR U SC Columbia U VT	ND SU U CA River U MA Amherst VA CommonWeal	Auburn U IN U Bloom U Central FL U DE WV U
2	Temple U U Toledo U TX El Paso	S Ill U Carbon U OR	Geo Mason U AL Huntsville U ID U MD Balt Co U MS U TX Dallas	Fl Intl MI Tech U TX Tech U Houston U LA Lafayette U ME U MT U ND U RI U WY	**Binghamton** GA SU NJ IT Old Dom San Diego SU Wichita SU
1		CO schl of Mines SD SU U Al	Jackson SU Will & Mary	Kent SU MO U of S&T NC A&T U Memphis U NV Las Vegas U So MS Wright State U	Bowling Gr Fl Atlantic Miami Ox N Ill U Northern AZ U OH U U Akron U MO Kansas City U MO St Louis U NC Greensboro U North TX U TX Arlington U WI Milwaukee West MI U

Percent of Operating Budget State Supported

In addition to showing where each institution falls on these distributions, we have chosen to highlight our SUNY sister university centers in red, as they are a natural peer group and often referenced in most internal benchmarking exercises. Other typologies developed for initial information sharing might include juxtaposing enrollment by the percentage of students receiving Pell grants, or student selectivity with other financial resource measures like percentage of budget allocated to instruction or student support services.

Examining these institutional typologies is also informative for the analysts. Deans and senior staff often have “insider insights” into other institutions and their cultures owing to either personal experience or professional collaborations/interactions with their counterparts across the country. This can help inform variable and measure selection, as well as aid in adding additional context to the eventual result set.

Once the institutional landscape is satisfactorily explored and campus leaders are familiarized with the data typologies, we then turn to statistical techniques to begin selecting peer institutions. The rank distance method used by Berthold (1996) is very similar to that used by the University of Kansas and described by Teeter and Christall (1987), but uses a percentile rank order of institutions on each measure rather than Z scores to calculate similarity/dissimilarity. A second method, cluster analysis, utilizes principal components analysis and factor scores to group universities that are determined to be similar across specified dimensions into clusters. This technique was developed by Terenzini et al. (1980), who were amongst this technique’s early pioneers in the late 1970s.

The Kansas classification described by Teeter and Christall (1987) utilized a weighting scheme to elevate the importance of certain variables after standardization. While this analysis does not use explicit weights for the variables chosen, an implicit weighting scheme is active in that seventeen financial, twelve size, twenty-five complexity, and five quality measures are used in the analysis. Hence, elements of finance and size are more dominant in assessing institutional similarity/dissimilarity with the target institution. In addition, it should be noted that many of the measures used are highly correlated with each other, so they in effect may very well be measuring the same variable or dimension. When we reduce the number of measures used in the analysis by about half, and with an eye toward equalizing the number of measures across the dimensions of financials, size, complexity and quality, we find that where institutions fall out in proximity to the target institution changes as well. Table 1 below shows Albany’s top 20 neighbors, and selected metrics, when an initial rank-distance analysis run is used with the seventy-eight variables initially identified for exploration.

Table 1 – Initial Extraction using the Rank-Distance method

		Financials				Size	Complexity					Quality		
	Composite Rank	Rnk - total_expenses	Rnk - tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	Rnk - tuit_fee_state_app_per_stu_FTE	Rnk - Avg_sal_FT_prof_9 montheq uated	Rnk - stu_FTE_enroll	Rnk - pct_Fac_FT	Rnk - stu_pct_min	Rnk - pct_stu_intl	Rnk - deg_bach_pct_psych	Rnk - deg_bach_pct_soc_sci	Rnk - sat_mid	Rnk - frosh_admit_rate	Rnk - frosh_yield_rate
UALBANY	**0**	**67**	**36**	**65**	**110**	**35**	**6**	**75**	**76**	**130**	**130**	**61**	**23.5**	**10**
U CA Santa Cr	**2,017**	53	50	63	90	40	35	116	3	128	119	79	79	6
Binghamton	**2,037**	34	108	56	106	25	25	79	128	134	112	126.5	9	14.5
U CA River	**2,101**	57	90	67	104	51	73	130	57	125	122	26	114.5	1.5
U NH	**2,227**	45	41	50	80	29	12	6	7	100	78	55	79	17
U MA Amherst	**2,239**	84	86	91	88	77	80	51	75	111	85	82	63.5	8
Stony Brook	**2,239**	114	22	128	127	65	24	103	125	133	109	108	9	16
U CA Santa B	**2,315**	77	93	92	118	71	88	110	26	115	124	104.5	20	3.5
U Buffalo	**2,331**	89	133	119	123	87	21	59	133	118	102	81	27.5	25.5
U CT	**2,354**	110	49	133	128	73	117	61	82	112	105	112	23.5	19.5
U MD Balt Co	**2,367**	19	73	41	76	16	42	109	93	131	120	106.5	70.5	50.5
Geo Mason	**2,374**	59	79	28	111	75	49	95	66	96	98	78	54.5	25.5
U SC Columbia	**2,426**	73	70	44	73	85	60	64	42	66	68	100.5	39.5	66
U OR	**2,444**	66	39	42	48	62	30	31	88	93	133	56.5	118	32.5
U HI Manoa	**2,478**	85	28	100	55	42	55	129	110	72	86	48.5	63.5	66
Rutgers New Br	**2,554**	111	101	130	129	115	31	111	85	126	121	100.5	48	32.5
U OK Nornam	**2,558**	75	52	59	69	66	81	77	77	27	73	87.5	97.5	92.5
U CA Irvine	**2,564**	113	15	85	117	95	57	128	68	129	132	92	20	10
U IL Chicago	**2,645**	118	10	113	100	79	11	115	103	132	64	61	54.5	36.5
U TX Dallas	**2,653**	28	58	68	122	21	41	106	130	116	59	118	27.5	101
U KS	**2,654**	96	64	89	85.5	92	58	29	94	91	89	77	127	92.5

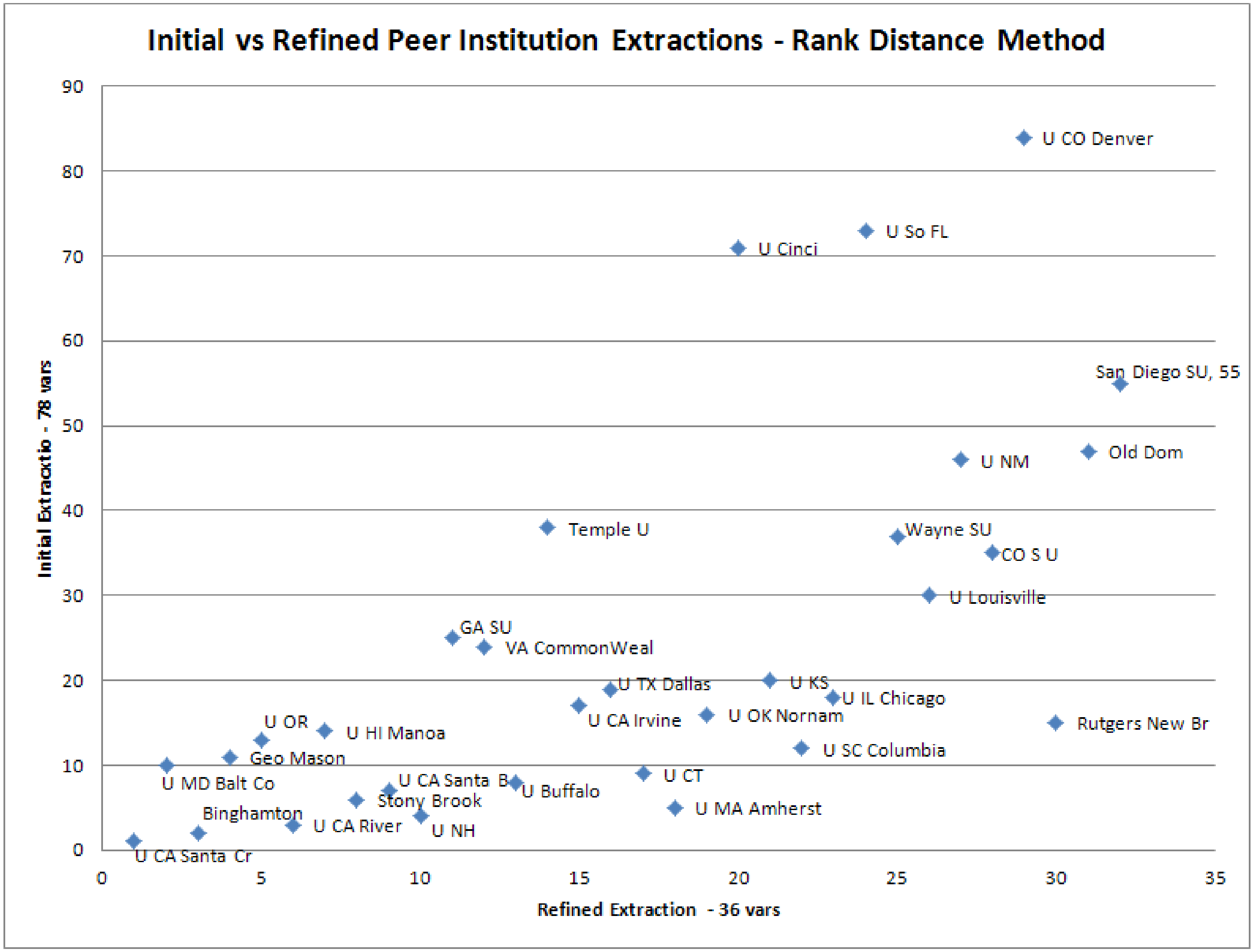
Due to space and visual presentation limitations, Table 1 shows only five of the seventeen Financials measures used; one of the twelve size measures, five of the 25 complexity measures, and three of the quality measures used. It is worth noting that several of institutions on Albany's existing peer institution list are present in this new ordering.

In order to explore the possible impact of collinearity among measures, as well as redundancy and attendant issues around inherent weighting by virtue of the differing number of measures in each dimension, we re-ran the rank-distance analysis using nine financials measures, eight in the size dimension, reduced the number of measures in the complexity dimension to fourteen, and held the quality dimension to its existing five measures. Although similar, there is enough variation in the resulting peer set to give one pause about relying solely on this method of extracting peer institutions. Table 2 shows the new institutional rank ordering, and includes the new top 20's rank prior rank on the initial extraction.

Table 2 – Refined Extraction using the Rank-Distance method

				Financials			Size	Complexity					Quality		
Inst_Abbrev	Initial Order	Refined Order	Composite Rank	Rnk - tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	Rnk - tuit_fee_state_app_per_stu_FTE	Rnk - Avg_sal_FT_prof_9 montheq uated	Rnk - stu_FTE_enroll	Rnk - pct_Fac_FT	Rnk - stu_pct_min	Rnk - pct_stu_intl	Rnk - deg_bach_pct_psych	Rnk - deg_bach_pct_soc_sci	Rnk - sat_mid	Rnk - frosh_admit_rate	Rnk - frosh_yield_rate
UALBANY			**0**	**36**	**65**	**110**	**35**	**6**	**75**	**76**	**130**	**130**	**61**	**23.5**	**10**
U CA Santa Cr	1	1	**992**	50	63	90	40	35	116	3	128	119	79	79	6
U MD Balt Co	10	2	**1,052**	73	41	76	16	42	109	93	131	120	106.5	70.5	50.5
Binghamton	2	3	**1,054**	108	56	106	25	25	79	128	134	112	126.5	9	14.5
Geo Mason	11	4	**1,077**	79	28	111	75	49	95	66	96	98	78	54.5	25.5
U OR	13	5	**1,106**	39	42	48	62	30	31	88	93	133	56.5	118	32.5
U CA River	3	6	**1,124**	90	67	104	51	73	130	57	125	122	26	114.5	1.5
U HI Manoa	14	7	**1,140**	28	100	55	42	55	129	110	72	86	48.5	63.5	66
Stony Brook	6	8	**1,153**	22	128	127	65	24	103	125	133	109	108	9	16
U CA Santa B	7	9	**1,178**	93	92	118	71	88	110	26	115	124	104.5	20	3.5
U NH	4	10	**1,201**	41	50	80	29	12	6	7	100	78	55	79	17
GA SU	25	11	**1,205**	116	33	91	84	126	121	52	105	100	43	11	133
VA CommonWe	24	12	**1,209**	84	46	67	98	20	101	53	114	29	48.5	73	44.5
U Buffalo	8	13	**1,246**	133	119	123	87	21	59	133	118	102	81	27.5	25.5
Temple U	38	14	**1,276**	14	55	116	113	4	91	41	65	42	56.5	48	50.5
U CA Irvine	17	15	**1,281**	15	85	117	95	57	128	68	129	132	92	20	10
U TX Dallas	19	16	**1,283**	58	68	122	21	41	106	130	116	59	118	27.5	101
U CT	9	17	**1,285**	49	133	128	73	117	61	82	112	105	112	23.5	19.5
U MA Amherst	5	18	**1,305**	86	91	88	77	80	51	75	111	85	82	63.5	8
U OK Nornam	16	19	**1,310**	52	59	69	66	81	77	77	27	73	87.5	97.5	92.5
U Cinci	71	20	**1,329**	72	79	52	96	3	42	79	33	33	69.5	48	44.5

In the table above we see that some institutions now enter the top 20 after being previously found to be as far away as overall rank 71. Conversely, institutions formerly in the top 20 ended up being as far from Albany as 30th. The following diagram graphically illustrates the intersection of these two rank-distance method runs. Taken together, these tables and the graphical illustration are perhaps cause for concern. By using a method that appears to be very sensitive to the number and types of measures used, and with potential issues of redundancy, stakeholder confidence in the eventual peer list might very well be compromised. While a weighting scheme might alleviate some of these concerns, they do not address the redundancy issue, and would add to the complexity of the analysis, and its communication to stakeholders and interested parties.



Initial vs Refined Peer Institution Extractions - Rank Distance Method
Initial Extracktio - 78 vars
Refined Extraction - 36 vars
0
10
20
30
40
50
60
70
80
90
0
5
10
15
20
25
30
35
U CO Denver
U So FL
U Cinci
San Diego SU, 55
U NM
Old Dom
Temple U
Wayne SU
CO S U
U Louisville
GA SU
VA CommonWeal
U TX Dallas
U KS
U IL Chicago
U CA Irvine
U OK Nornam
Rutgers New Br
U OR
U HI Manoa
Geo Mason
U SC Columbia
U MD Balt Co
U CA Santa B
U Buffalo
U CT
Stony Brook
U MA Amherst
Binghamton
U CA River
U NH
U CA Santa Cr

To address these concerns, we employ factor analysis (varimax rotation, and Kaiser's criterion for Eigenvalue selection) and a cluster analysis technique (complete linkage, hierarchical aggolomerative). The factor analysis uses principle components analysis to reduce the original seventy-eight measures to fourteen factors that reflect institutional dimensions of import. By definition, principle components analysis partitions the observed variance in the measures to unique factors that are completely uncorrelated. The researcher can though influence factor composition by choosing the extraction technique, and in deciding where to halt factor formation, so some subjectivity is still present, albeit in a limited manner. The dimensions that surfaced in this analysis are characterized as reflecting: Size; Wealth; Dedication of Resources to Mission; Undergraduate Environment; Undergraduate Program Complexity; Student to Faculty Ratio; Diversity of the Student Body; Graduate Program Complexity; Full-time Faculty Emphasis; Liberal Arts Programming; Engineering-Medical Emphasis; Institutional Commitment to Student Aid; Math Programmatic Emphasis at the Graduate Level, and Federal R&D Activity. The factor scores are exported and saved in the dataset, and are then used in a cluster analysis. Institutional clustering via a dendogram diagram can then be examined to determine how institutions cluster with the target institution.

Each of the factors receives equal weight in building the clusters, and complete data is required for each institution. Initially, a conscious decision was made not to weight the factors, even though no design induced implicit weighting scheme existed. After the first pass through, it was decided to deploy a weighting scheme as the initial clustering around the target institution was less than robust. The following weights were used for this exploratory analysis: Wealth 15%; Size, Dedication of Resources to Mission, UG Environment, Student Faculty Ratio, Diversity of Student Body, FT Faculty Emphasis, and Liberal Arts Education 10%; UG Program Complexity and Graduate Program Complexity 5%; Engineering – Medical 2%; Institutional Student Aid, Math (graduate) Emphasis, and Federal R&D 1%. These relative weights were assigned by the analysts for this particular research effort, but the importance of vetting and assigning weights with input from the campus decision makers who will be relying on the resulting peer lists cannot be understated.

The cluster analysis calculates a standard Euclidean distance measure for each institution based on the standardized factor scores, and uses the "complete linkage" hierarchical agglomerative technique to group institutions into relatively homogeneous clusters. The clusters are formed based on the minimum maximum distance score between institutions, which is compared at each successive step until the researcher decides to stop cluster formation (based on professional judgment about diminishing returns). Each institution's cluster can be as small as two campuses, or can be built larger to incorporate larger numbers, but at each successive step, the number of institutions that join clusters is variable, and depends on the clustering algorithm, the weights assigned to factors (or not) and the researcher's objective.

The underlying purpose of this analysis is to identify those ten to twenty institutions that group closest to the target institution. Institutions can be grouped by use of a dendogram, which traces the clustering pattern of any institution to successive institutions or groups of institutions. By definition, similarity between institutions becomes less distinct as the clusters incorporate come to incorporate additional schools. A dendogram is then produced based on the underlying weighted factors and the distance between institutions on them. Visual inspection is then used to determine where to stop the clustering process as more and more institutions join the initial cluster.

Results

Univariate statistics for the seventy-eight measures used are reported in Appendix A. Even though only public doctoral granting universities with *very high* or *high* research emphasis are included in the analysis, brief review of the means and standard deviations suggests a great deal of variability on many of the measures across these institutions. This further reinforces the need for a comparison strategy. As in our prior explorations (1989, 1996, 2006), the financial measures exhibited the most variability, followed by the size and then quality measures. There was considerably less variation in the complexity measures, suggesting either that this concept may be more difficult to measure or that these universities are indeed more alike in this respect.

The (truncated) bivariate results reported in the correlation matrix in Appendix B indicate measures, in general, that are highly correlated with each other, and reinforces the concerns noted above around redundancy of measures. For example, total revenues has a Pearson's correlation coefficient above 0.50 with every financial measure except for percent of operating revenue devoted to research expenditures (r = -.137) and average associate professor's salary (r = 0.427). Furthermore, it is also highly correlated with freshman retention and average SAT scores, degree production, and the number of full-time faculty. Other measures were also highly correlated both within and across the hypothetical dimensions of funding, size, quality, and complexity. While more in-depth discussion of these relationships is not entertained here due to space limitations, the important implication is that a factor analytic technique that controls for multicolinearity would be more suitable to this data set. The rank distance method is though seen as beneficial in terms of educating campus decision makers about relevant data its spread among institutions.

The second method used to develop a set of peer institutions is a factor and cluster analysis technique. When the seventy-eight measures described above are subjected to principal components analysis, with varimax rotation and using Kaiser's criterion for Eigenvalue selection, fourteen factors emerge which explain seventy-seven percent of total variance in the data set. Table 3 shows the factor loadings and the resulting dimensions.

These dimensions suggest a further delineation of the four general dimensions hypothesized earlier. The size dimension explains the most variance (20.1%) of the fourteen extracted factors. Wealth and dedication of resources to mission functions explain just over eight percent of variance each. The undergraduate environment and undergraduate program mix explain just over six percent of variance each. The remaining factors explain five and fewer percent of total variance, culminating in total explained variance across all factors of 77.5 percent. The factor loadings show that the previous notion of quality described earlier can be better defined as undergraduate quality, as the six-year graduation rate, freshmen retention rate, percent of undergraduates living in dorms, and the freshmen acceptance rate are particularly relevant to the undergraduate experience. Consistent with Szelest's (1996) earlier analysis, it the student faculty ratio does not load heavily with any other measures, but by itself, it does explain 5.1 percent of the total variance. Interestingly, the state supported portion of the operating budget has a negative loading on the size dimension, but state support per student FTE loads positively. The implication of this might be that larger institutions are overall less dependent on state support than their smaller counterparts, but on a per student basis, state's fund larger institutions better than smaller ones.

As noted earlier, after an initial exploratory cluster analysis was conducted without factor weights, it was decided that using weights might provide a more robust clustering solution. The following weights were assigned: Wealth 15%; Size, Dedication of Resources to Mission, UG Environment, Student Faculty Ratio, Diversity of Student Body, FT Faculty Emphasis, and Liberal Arts Education 10%; UG Program Complexity and Graduate Program Complexity 5%; Engineering – Medical 2%; Institutional Student Aid, Math (graduate) Emphasis, and Federal R&D 1%. Again, these relative weights were assigned by the analysts for this particular research effort. Future efforts should include input from campus decision makers.

Appendix D illustrates the dendogram that graphically depicts the institutional clustering sequence that results when the factor scores are submitted to the clustering algorithm. Of primary interest in this exercise are those institutions that cluster with Albany, the target institution. Examining the dendogram shows that Albany first clusters with University of Maryland – Baltimore County. These two institutions are then joined by the University of Colorado – Boulder and the University of Oregon, both of which were prior Albany peers. At the next iteration, George Mason University joins this small cluster of four institutions, followed by a larger cluster composed of the University of Maryland - College Park, Rutgers New Brunswick, the University of Massachusetts – Amherst, and the University of Connecticut. Next, SUNY Buffalo, SUNY Stony Brook, and SUNY Binghamton join the cluster. Then the Georgia Institute of Technology and Michigan Technological University join, followed by the New Jersey Institute of Technology and the University of Texas – Dallas. The Colorado School of Mines joins the cluster by itself, and finally, Pennsylvania State University, the University of Pittsburgh, Temple University, and the University of Delaware join the cluster. At this point, the next set of institutions to join the cluster would be all the remaining institutions, as they form one large block at this stage, rather than additional smaller groupings. This appears to be a natural stopping point.

Now that a cluster of peer institutions has been formed, we might ask "how it might best be described?" Seven out of these twenty institutions have hospitals or medical schools (U CT, U MD College Park, SUNY Buffalo and Stony Brook, Temple U, U Pittsburgh, and Penn State U) while Albany does not. Six of the potential peers have land grant status, which interestingly include the entire sub-cluster of U MD College Park, Rutgers New Brunswick, U MA Amherst, and U CT, and then two institutions from the very last cluster to load with Albany, Penn State U and the University of Delaware. Table 3 displays the original percentile rank measures for the peer cluster. One could also examine institutional rank on the factor scores themselves.

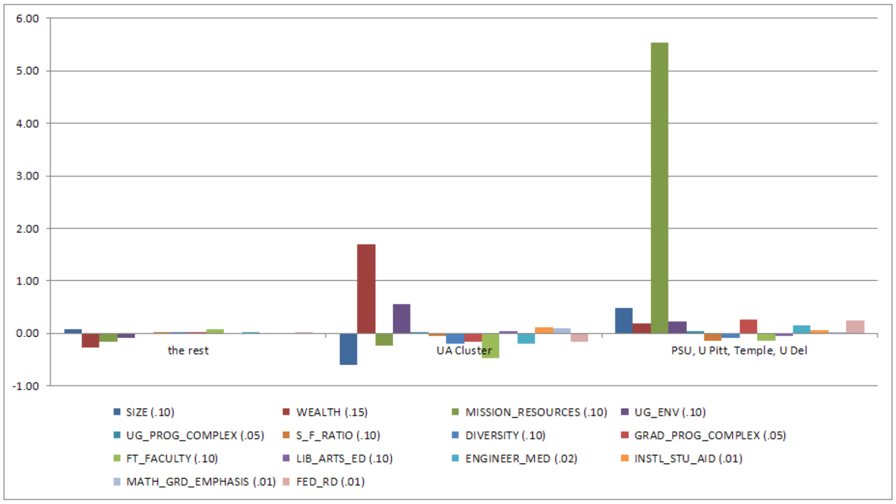
In Table 3 we see that the first institution to load in the cluster with the target institution is UMD Baltimore County. Ranks on several of the measures depicted appear relatively close, so one can see why this institution might load as the first peer institution. As we work our way down the table, much more distance between measure rankings is observed, so making the case for inclusion of institutions into the final peer group becomes less firm.

Table 3 – Rank on Individual Metrics for Cluster Peers

	Financials			Size	Complexity					Quality		
Inst_Abbrev	Rnk - tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	Rnk - tuit_fee_state_app_per_stu_FTE	Rnk - Avg_sal_FT_prof_9 montheq uated	Rnk - stu_FTE_enroll	Rnk - pct_Fac_FT	Rnk - stu_pct_min	Rnk - pct_stu_intl	Rnk - deg_bach_pct_psych	Rnk - deg_bach_pct_soc_sci	Rnk - sat_mid	Rnk - frosh_admit_rate	Rnk - frosh_yield_rate
UALBANY	**36**	**65**	**110**	**35**	**6**	**75**	**76**	**130**	**130**	**61**	**23.5**	**10**
U MD Balt Co	73	41	76	16	42	109	93	131	120	106.5	70.5	50.5
U OR	39	42	48	62	30	31	88	93	133	56.5	118	32.5
U CO Boulder	29	38	84	107	27	41	34	110	97	92	114.5	36.5
Geo Mason	79	28	111	75	49	95	66	96	98	78	54.5	25.5
U CT	49	133	128	73	117	61	82	112	105	112	23.5	19.5
U MA Amherst	86	91	88	77	80	51	75	111	85	82	63.5	8
Rutgers New Br	101	130	129	115	31	111	85	126	121	100.5	48	32.5
U MD College Pk	77	114	120	114	51	94	108	48	127	126.5	12.5	39.5
Binghamton	108	56	106	25	25	79	128	134	112	126.5	9	14.5
Stony Brook	22	128	127	65	24	103	125	133	109	108	9	16
U Buffalo	133	119	123	87	21	59	133	118	102	81	27.5	25.5
GA T	23	105	126	49	129	85	134	3	4	130	48	82
MI Tech U	104	71	32	4	98	4	126	4	3	114	90.5	44.5
U TX Dallas	58	68	122	21	41	106	130	116	59	118	27.5	101
NJ IT	125	102	134	5	16	119	132	1.5	1	63	20	101
CO schl of Mines	34	45	109	1	34	26	121	1.5	6	122.5	48	6
Temple U	14	55	116	113	4	91	41	65	42	56.5	48	50.5
U Pittsburgh	99	115	112	86	75	36	96	122	94	119	31.5	21.5
U DE	134	104	121	53	127	44	87	38	99	98	37	21.5
PA SU	51	134	114	129	101	37	113	35	31	102.5	26	44.5

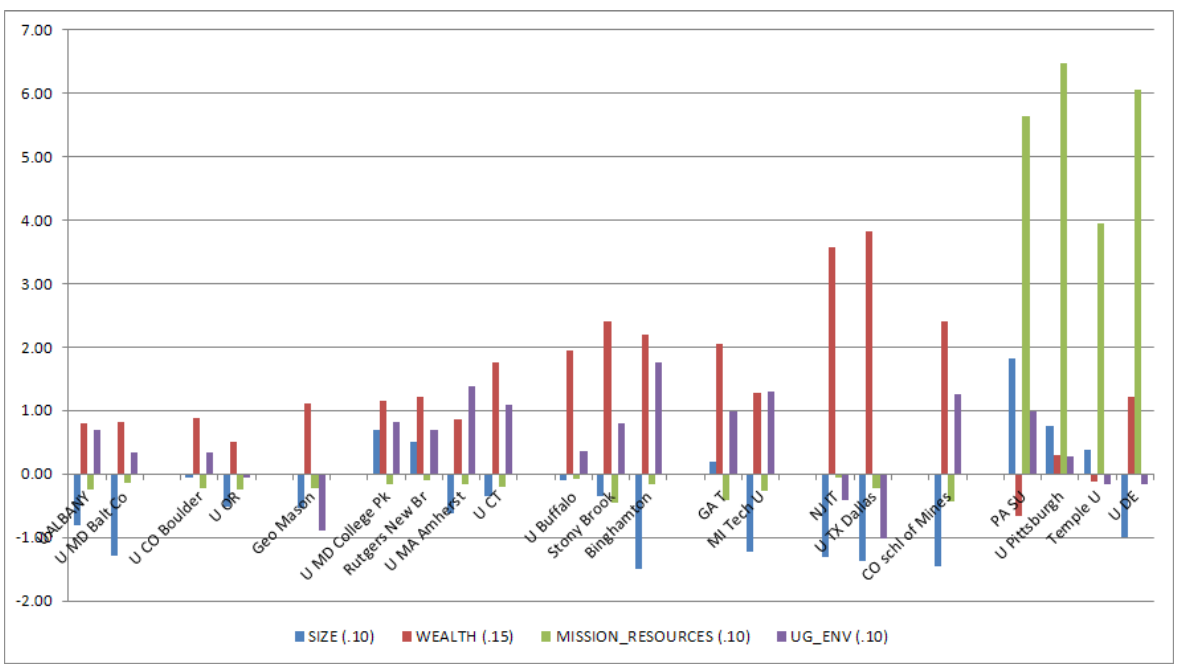
Alternative ways of looking at the inherent attributes of the institutions in the peer clusters is to graph the factor scores. Chart 1 below shows the mean factor scores for all of the institutions that would have loaded next with the final Albany cluster, The Albany (UA) cluster less the final four institutions that loaded with Albany, and then the final four institutions that loaded into the final cluster (Penn State U, U Pittsburgh, Temple U, and U DE). Quick visual inspection shows that the Albany cluster is indeed different from the other clusters on these dimensions. It should be noted that the Penn State U, U Pittsburgh, Temple U, and U DE cluster is part of the final Albany cluster, but has been separated from it for illustration purposes. Indeed, any one of the iteratively added sub-clusters could be separated out to allow for closer inspection of how it reflects the factor dimensions.

Chart 1 – Mean factor scores for three clusters of institutions



An alternative to looking at cluster differences to build confidence in the peer group candidates is to look at the attributes of the institutions that are in the cluster and highlight their similarities. Chart 2 below shows the mean factor scores for the twenty-one institutions in the final Albany peer cluster, grouped by their sub-clusters. Only the first four dimensions, of size, wealth, mission resources, and undergraduate environment, are shown due to space limitations.

Chart 2 – Within Cluster Attributes



It is important to keep in mind that the factors scores were weighted in the final solution. Visual inspection seems to confirm that, as move from left to right on the chart, that the candidate institutions are similar to a large extent on these measures. While we observe less similarity as we get to the final four institutions at far right, it should be noted that they nonetheless are more like the institutions at left, across all of the dimensions, than they are like all of the other institutions which are not members of this cluster. Whether or not to include these final four institutions in the final peer group that will constitute Albany's "official peers" is really a political decision that needs to be made by campus leaders.

Discussion

In summary, the rank distance and factor and cluster analysis techniques provided an overlapping group of peer institutions for the target institution, SUNY Albany. In both analyses, campus decision makers have opportunity to play an integral role in developing the list of variables and their respective measures used to differentiate institutions. With both techniques, the institutional research staff, as well as higher level campus decision makers, can clearly see how Albany compares to other institutions across financial, size, administrative complexity, graduate study emphasis and quality measures. Increased familiarity with and knowledge about these aspects of our own institution and of our chosen peers is of importance as we move forward with implementing our new strategic plan. Benchmarking to external institutions is often requested, and having an up-to-date set of peers adds confidence to the exercise.

There are limitations to these methods. The rank distance method employed did not assign weights to the individual measures. While an implicit weighting scheme was in effect due to the different number of measures used to represent finance, size, quality, and complexity variables, arguments can certainly be made to alter the variable weighting, or the number of measures. As demonstrated, the number of measures used can indeed have an impact on the result set. In the past, our analyses were guided by a desire to seek peer institutions more closely related on the financial and size variables, and this analysis was conducted similarly. We will need to revisit this approach as we work with campus leaders to develop a formal set of peer institutions.

Another limitation of the rank distance method is that basing the distance measure on percentile ranks instead of on factor or standardized scores distorts the magnitudes of the differences between institutions. The percentile rank approach creates a distance measure based on the <u>rank order of institutions</u> for different measures rather than upon the magnitude of their differences from each other. The actual distances in the raw data may be greater or smaller than the distance between ranks.

A final limitation of this methodology worth noting is that many of the measures used, as noted, are highly correlated with each other. In other words, to a certain degree, they may be measuring the same concepts. While the measures used address different concerns campus decision makers harbored, the end result may be confounded by the highly related nature of the measures.

The factor and clustering technique, which alleviates the problem of highly correlated measures by factoring them into fourteen completely uncorrelated dimensions has an advantage in this respect. In this analysis, we decided to use weights on each of the fourteen dimensions to provide for a more robust clustering solution.

A well noted limitation of factor analysis is that the factors (dimensions) used for clustering are more difficult to assess. This is particularly true with respect to the percentile ranks used in the first analysis. While standardized factor scores do not easily lend themselves to meaningful interpretation, once can still compare factor scores across institutions, or by cluster, to convey a sense of similarity/dissimilarity.

Both of these methods provide a means of identifying peer institutions. Neither should be viewed as a turn-key approach. The rationale for the undertaking in the first place plays a significant role in choosing variables and measures of them for consideration. Acceptance of a peer institution group by campus decision makers and relevant external audiences largely depends on accommodating and incorporating their intuition, concerns, and political objectives. The peer review process is a learning process. Through it we learn not only about the funding, quality, or size of other institutions, but more importantly, we learn about how our own institution stands in those respects.

Future Research

Future research might address the methodological limitations noted above. These include important methodological aspects, such as the sources of variables and selection of variables --- that in turn strongly influence the composition of the eventual peer groupings.

Due to resource, accessibility, and other constraints, this study has focused only on data available through the IPEDS database. However, there are other reliable sources of institutional data from sources such as the National Research Council (NRC), the American Association of University Professors (AAUP), US News & World Report, and the National Science Foundation (NSF) which can provide additional measures for peer analyses. Looking to the future, these sources can be used to incorporate variables that better measure, for example, the quality dimension through faculty work products such as the number of publications per faculty or the number of citations per faculty. The financial dimension could be enriched by evaluation of NSF funded Research and Development expenditures at the programmatic level. Other possibilities may exist as well.

References:

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis. Beverly Hills, CA: Sage

Archer, S. J., Reiss E., Armacost R. L., Sun Y., & Fu, Y. (2009). *A Large Metropolitan Research University's Approach to Benchmarking Peer Selection.* Proceeding of Association for Institutional Research 49th Annual Conference, Atlanta, Georgia.

Berthold, C. A. (1996). *Aids to Navigation: Using Interinstitutional Databases in the University of Maryland System*. Paper presented at the 36th Annual Forum of the Association for Institutional Research, Albuquerque, NM, May 1996.

Brinkman, P. T. and Teeter, D. J. (1987). Methods for Selecting Comparison Groups, in *Conducting Interinstitutional Comparisons*, Paul T. Brinkman (ed), New Directions for Institutional Research, no. 53., Vol XIV, Number 1., San Francisco: Jossey-Bass.

Gaylor, S.S. (2009). *Use of a Comparison Group and Modeling to Inform Decision-Making and Planning.* Proceeding of Association for Institutional Research 49th Annual Conference, Atlanta, Georgia.

Graham, H.D. & Diamond, N. (1996). *The Rise of American Research Universities: Elites and Challengers in the Postwar Era.* Baltimore: Johns Hopkins University Press.

Gnanadesikan, R., Kettenring, J. R., & Tsao, L., (1995). Weighting and selection of variables cluster analysis. *Journal of classification*, 1995, 12(1), 113-136.

Gater, D.S., (2003). *Using National Data in University Rankings and Comparisons. The Center Report.* Proceeding of Association for Institutional Research 43rd Annual Conference, Tampa, Florida.

Gater, D. S., & Lombardi, J. V., (2001). *The Use of IPEDS/AAUP Faculty Data in Institutional Peer Comparisons. The Center* Report. Proceeding of Association of American Universities Annual Meeting, Columbus, Ohio.

Ingram, J. A. (1995). Using IPEDS Data for Selecting Peer Institutions. (ERIC Document Reproduction Service No. ED387010).

IU Northwest Peer Institution Selection and Analysis (2006) Report presented by the Board of Trustees.

Lang, D. W. (2000). Similarities and Differences: A Case Study in Measuring Diversity and Selecting Peers in Higher Education. *Higher Education 39*: 93–129, 2000. Kluwer Academic Publishers. Printed in the Netherlands.

Lang, D. W., & and Zha, Q., (2004). Comparing Universities: A Case Study between Canada and China. *Higher Education Policy*, 2004, 17, (339–354).

National Research Council (1995), *Research-Doctorate Programs in the United States, Continuity and Change*, Goldberger, M.L., Maher, B.A., Flattau, P.E., (eds), Washington, D.C.: National Academy Press.

Nzeukou, M., & Muntal, D. (2010). *A Knowledge-Based Selection Methodology of Peer Institutions.* Proceeding of Association for Institutional Research 50th Annual Conference, Chicago, Illinois.

Panayotova, E. (2009). *Selecting Peers in Higher Education-Opportunities and Challenges in the Context of Institutional Growth.* Proceeding of Association for Institutional Research 49th Annual Conference, Atlanta, Georgia.

Szelest, B. P. (1996). *In Search of Peer Institutions: Two Methods of Exploring and Determining Peer Institutions.* Proceeding of North East Association for Institutional Research 23rd Annual Conference, Princeton, New Jersey.

Teeter, D. J. & Brinkman, P. T. (1992). Peer Institutions in *The Primer for Institutional Research*, Whiteley, M.A., Porter, J.D., and Fenske, R.H. (eds), Tallahassee: Association for Institutional Research.

Teeter, D. J. & Christal, M. E. (1987.) Establishing Peer Groups: A Comparison of Methodologies. *Planning for Higher Education*, Vol. 15, No. 2, 1987.

Terenzini, P.T., Hartmark, L., Lorang, W.G. Jr., & Shirley, R.C.. (1980). A Conceptual and Methodological Approach to the Identification of Peer Institutions. *Research in Higher Education,* Vol. 12, No. 4, 1980, Agathon Press, Inc.

Trainer , J. F. (2008) The Role of Institutional Research in Conducting Comparative Analysis of Peers .*New Directions for Higher Education, no. 141, Spring 2008.*

Volkwein, J.F. and Malik, S.M. (1996). *State Regulation and Administrative Flexibility at Public Universities*. Paper presented at the 1996 AIR Forum in Albuquerque, NM, and accepted for publication in *Research in Higher Education*, Vol. 38, No. 1, February 1997.

Weeks, S. F., Puckett, D., & Daron, R. (2000). Developing peer groups for the Oregon University System: From politics to analysis (and back). *Research in Higher Education, 41,* 1–20.

Xu, J. (2008).Using the IPEDS Peer Analysis System in Peer Group Selection*. Professional File, Association for Institutional Research,* Number 110, Winter 2008.

Zhao, J., & Dean, D. C. (1997). *Selecting Peer Institutions: A Hybrid Approach.* Paper presented at the
1997 AIR Forum in Orlando, FL

Appendix A – Univariate Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
land_grant	134	0	1	0	0
engineering_presence	134	0	1	0.93	0.26
hosp_meddeg	134	0	1	0.48	0.50
total_expenses	134	48,370,037	5,098,288,000	983,075,545	896,802,055
tot_revenues	134	180,701,338	5,786,349,000	1,123,555,156	1,009,646,001
tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	134	0.14	1.02	0.44	0.13
tuit_fee_state_app_per_stu_FT	134	8,694	34,550	17,326	4,891
expend_research	134	4,082,478	862,386,314	194,475,235	193,798,298
pct_of_op_exp_research	134	0.01	5.16	0.27	0.56
Tot_SE_RD	134	8,396	1,007,198	231,264	224,085
Fed_SE_RD	134	4,285	636,216	122,988	130,018
Pct_SE_RD_Federal	134	0.19	0.89	0.53	0.15
pct_of_op_exp_instruct	134	0.11	6.51	0.45	0.79
pct_of_op_exp_pub_serv	134	0.00	0.95	0.07	0.11
pct_of_op_exp_acad_supt	134	0.03	1.62	0.13	0.24
pct_of_op_exp_stu_serv	134	0.01	0.94	0.07	0.12
pct_of_op_exp_instl_support	134	0.01	1.35	0.11	0.20
endow	134	359,183	5,914,285,000	435,486,621	714,920,696
Avg_sal_FT_prof_9monthequat	134	72,062	158,660	112,165	16,064
Avg_sal_FT_assocProf_9moeq	134	61,870	115,440	80,264	8,684
avg_sal_FT_assntProf_9moequ	134	54,253	92,433	69,005	7,524
pct_ftft_ug_rec_st_loc_grnt_aid	134	1.00	96.00	34.61	23.98
pct_ftft_ug_rec_instl_grnt_aid	134	6.00	93.00	48.08	16.96
stu_FTE_enroll	134	4,659	59,850	23,044	10,517
stu_ft_UG_enroll	134	3,474	45,597	17,244	7,912
stu_pt_UG_enroll	134	76	12,234	2,683	2,241
stu_ft_GRAD_enroll	134	444	13,531	3,847	2,754
stu_pt_GRAD_enroll	134	74	9,556	2,408	1,500
ftft_frosh	134	494	8,261	3,497	1,637
ft_transfers	134	0	4,178	1,218	701
GT_FT_faculty	134	312	5,693	1,622	970
Tot_N_FT_nonFAC	134	0	21,178	3,492	3,266
degs_bach	134	473	11,810	4,089	2,202
degs_masters	134	202	3,914	1,336	766
degs_doct_resch	134	21	891	245	207
degs_doct_profess	134	0	1,356	222	245
pct_ftft_ug_rec_Pell	134	9.00	78.00	27.61	11.29
Pct_enroll_graduate	134	0.09	0.44	0.23	0.07
pct_ug_18_to_24	134	46.00	97.00	83.90	10.70
pct_Fac_FT	134	0.46	0.99	0.79	0.11
Pct_tot_employees_FT_fac	134	0.14	0.67	0.31	0.06
stu_pct_min	134	0.05	0.94	0.25	0.17
pct_stu_intl	134	0.00	0.19	0.06	0.03
dorm_cap_per_stu	134	0	1	0	0
deg_bach_pct_educ	134	0.00	0.20	0.06	0.05
deg_bach_pct_engineer	134	0.00	0.88	0.10	0.11
deg_bach_pct_eng_libarts_vis_arts	134	0.00	0.29	0.10	0.05
deg_bach_pct_bio	134	0.00	0.22	0.07	0.03
deg_bach_pct_math	134	0.00	0.07	0.01	0.01
deg_bach_pct_phil	134	0.00	0.04	0.01	0.01
deg_bach_pct_phys_sci	134	0.00	0.06	0.02	0.01
deg_bach_pct_psych	134	0.00	0.14	0.06	0.03
deg_bach_pct_soc_sci	134	0.00	0.28	0.11	0.05
deg_bach_pct_bus	134	0.00	0.34	0.17	0.07
deg_bach_pct_hist	134	0.00	0.08	0.02	0.01
deg_doct_resch_pct_educ	134	0.00	0.47	0.14	0.11
deg_doct_resch_pct_engineer	134	0.00	0.63	0.16	0.13
deg_doct_resch_pct_eng_libarts_vis_arts	134	0.00	0.24	0.05	0.05
deg_doct_resch_pct_bio	134	0.00	0.65	0.13	0.10
deg_doct_resch_pct_math	134	0.00	0.18	0.04	0.03
deg_doct_resch_pct_phil	134	0.00	0.08	0.01	0.01
deg_doct_resch_pct_phys_sci	134	0.00	0.35	0.11	0.06
deg_doct_resch_pct_psych	134	0.00	0.26	0.07	0.06
deg_doct_resch_pct_soc_sci	134	0.00	0.31	0.07	0.05
deg_doct_resch_pct_bus	134	0.00	0.13	0.02	0.03
deg_doct_resch_pct_hist	134	0.00	0.07	0.02	0.02
sat_mid	134	910	1,350	1,135	93
frosh_admit_rate	134	21	99	67	17
frosh_yield_rate	134	17	72	40	11
frosh_retent_rate	134	65	97	83	8
grad_rate_6yr	134	21	93	61	15
stu_fac_ratio	134	11	31	18	3

Appendix B – (truncated) Correlation Matrix (a complete correlation matrix is available from the authors, upon request)

Pearson Correlations

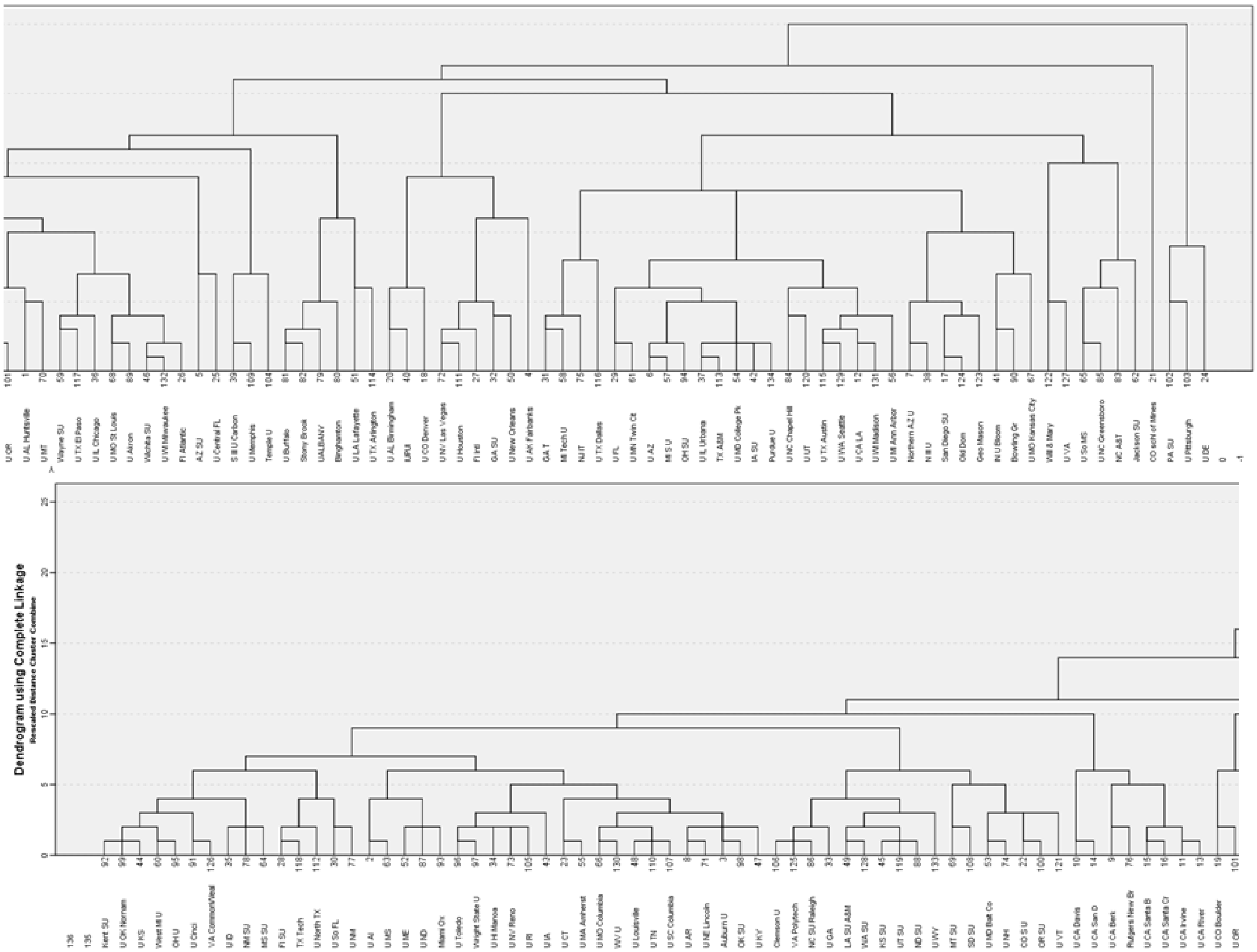
	land_grant	engineering_presence	hosp_meddeg	total_expenses	tot_revenues	tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	tuit_fee_state_app_per_stu_FTE	expend_research	pct_of_op_exp_research	Tot_SE_RD	Fed_SE_RD	Pct_SE_RD_Federal	pct_of_op_exp_instruct
land_grant	1	.226	.189	.094	.100	-.043	.323	.236	.093	.249	.130	-.185	.047
engineering_presence	.226	1	.215	.088	.098	-.077	.054	.127	.067	.107	.093	-.121	.037
hosp_meddeg	.189	.215	1	.492	.513	-.475	.407	.492	.115	.495	.481	.038	.033
total_expenses	.094	.088	.492	1	.938	-.658	.540	.788	-.137	.803	.771	.021	-.211
tot_revenues	.100	.098	.513	.938	1	-.662	.612	.844	.102	.845	.831	.063	.047
tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	-.043	-.077	-.475	-.658	-.662	1	-.286	-.565	.109	-.540	-.565	-.194	.286
tuit_fee_state_app_per_stu_FTE	.323	.054	.407	.540	.612	-.286	1	.583	.255	.572	.532	-.026	.173
expend_research	.236	.127	.492	.788	.844	-.565	.583	1	.301	.960	.945	.084	.097
pct_of_op_exp_research	.093	.067	.115	-.137	.102	.109	.255	.301	1	.261	.328	.167	.806
Tot_SE_RD	.249	.107	.495	.803	.845	-.540	.572	.960	.261	1	.955	.027	.074
Fed_SE_RD	.130	.093	.481	.771	.831	-.565	.532	.945	.328	.955	1	.233	.125
Pct_SE_RD_Federal	-.185	-.121	.038	.021	.063	-.194	-.026	.084	.167	.027	.233	1	.157
pct_of_op_exp_instruct	.047	.037	.033	-.211	.047	.286	.173	.097	.806	.074	.125	.157	1
pct_of_op_exp_pub_serv	.289	.066	.141	-.121	.043	.220	.187	.153	.759	.145	.153	.049	.851
pct_of_op_exp_acad_supt	-.007	.046	.104	-.193	.099	.152	.196	.157	.874	.119	.178	.135	.931
pct_of_op_exp_stu_serv	-.013	.048	.022	-.259	.021	.211	.129	.107	.901	.072	.139	.138	.899
pct_of_op_exp_instl_support	.011	.033	.010	-.242	.019	.272	.161	.091	.877	.064	.125	.166	.971
endow	.012	.010	.290	.653	.704	-.370	.404	.623	.202	.623	.682	.161	.115
Avg_sal_FT_prof_9monthequated	-.054	.047	.238	.531	.579	-.273	.587	.530	.183	.539	.520	.021	.184
Avg_sal_FT_assocProf_9moequated	-.025	.085	.218	.427	.468	-.233	.544	.436	.151	.444	.439	.047	.167

Appendix C – Factor Matrix

Rotated Component Matrixa

	Component													
	Size	Wealth	Dedication of Resources to Mission	UG Environment	UG Program Complexity	Student Faculty Ratio	Diveristy of Student Body	Graduate Program Complexity	FT Faculty Emphasis	Liberal Arts Education	Engineering - Medical	Institutional Student Aid	Math (graduate) Emphasis	Federal R&D
stu_ft_GRAD_enroll	.921	.234												
Tot_N_FT_nonFAC	.908													
degs_doct_resch	.907	.229		.158							-.100			
tot_revenues	.906	.171			.109	-.122					.140			
Tot_SE_RD	.878	.182		.125				-.170			.115			
expend_research	.871	.172		.101		-.102		-.174			.143			
total_expenses	.862	.189	-.271		.133	-.118					.141			
Fed_SE_RD	.842	.182	.127			-.128		-.182			.121			.189
GT_FT_faculty	.841	.126	.154				-.127	-.145		.169	.183			
stu_FTE_enroll	.840					.455					-.167			-.107
stu_ft_UG_enroll	.776			.125		.500					-.216	-.106		-.129
degs_bach	.762		.106	.120		.513					-.206			-.104
ftft_frosh	.749			.206		.415	-.121				-.210	-.187		-.115
degs_masters	.748	.243		-.191		.286			-.153	.184	-.129	.129	.260	
endow	.701	.136				-.226	-.199							.243
degs_doct_profess	.701			-.159	.140		-.141	.140	-.138	.165	.302		.150	
frosh_retent_rate	.555	.390		.537	.155				.129	-.122		.208		
tuit_and_fee_plus_state_approp_as_pct_of_tot_rev	-.521		.281			.203					-.510		.124	-.281
sat_mid	.485	.469		.426			-.315			-.152		.268		
tuit_fee_state_app_per_stu_FTE	.477	.359	.149	.246	.163	-.429			.175			-.101		-.226
Avg_sal_FT_assocProf_9moequated	.312	.833	.101	.157									.108	
avg_sal_FT_assntProf_9moequated	.420	.798		.140										
Avg_sal_FT_prof_9monthequated	.450	.756	.115	.206	.117									
deg_doct_resch_pct_educ	-.196	-.676									-.239	.240	.320	
pct_stu_intl	.266	.652			-.359	-.143					-.125			-.172
deg_bach_pct_educ	-.204	-.641		-.165	-.167	-.111		.171		.101	-.154	-.150	.172	-.196
deg_bach_pct_math		.539	-.120	.324		-.208		-.176	-.275	-.286		.150	-.177	
deg_doct_resch_pct_soc_sci	.111	.483			.430	.156			-.151		-.184		.221	-.142
Pct_enroll_graduate	.371	.434		-.193		-.259	-.115		-.244	.322	.250	.182	.424	
pct_of_op_exp_instl_support			.984											
pct_of_op_exp_acad_supt			.964											
pct_of_op_exp_stu_serv			.964											
pct_of_op_exp_instruct			.962											
pct_of_op_exp_research	.130		.905					-.112						
pct_of_op_exp_pub_serv			.847			-.101	-.153		.193					-.124
pct_ug_18_to_24	.287	.117		.818			-.210		.112				-.110	
dorm_cap_per_stu	-.102	.188		.781	.133	-.226			.148	-.104	-.112			
stu_pt_UG_enroll				-.712		.487	.110				-.101		.160	
grad_rate_6yr	.511	.279		.685	.140		-.214		.106					
frosh_yield_rate		-.130	-.108	-.597	-.229	-.109	-.182	.120	.383			.242	-.178	
deg_bach_pct_engineer		.377		.158	-.747	-.167		-.206		-.214				
deg_bach_pct_soc_sci	.282	.212		.208	.731			-.102		-.167		.137		.118
deg_bach_pct_psych		.267			.687		.248				.200	.103	.119	
deg_doct_resch_pct_engineer		.364			-.686		.158		.120	-.162		.134	-.128	.105
deg_bach_pct_hist	.137	.171		.281	.651	-.149						.303		.169
deg_bach_pct_phil	.140	.205		.188	.531				-.101			.211	-.200	.301
deg_bach_pct_bio	.329	.266	-.105	.276	.429		.414	-.226		-.147	.196			
stu_fac_ratio	-.105		-.139			.789								
ft_transfers	.385	.108		-.176	.179	.643	.180				-.198		.135	
deg_bach_pct_phys_sci	.139	.103		.195	.129	-.436	.104	-.277	-.117		-.182	.181	.106	.189
stu_pct_min		.147		-.144	.102	.104	.850					.152		
pct_ftft_ug_rec_Pell	-.239	-.211	-.112	-.325			.756		-.203					
deg_doct_resch_pct_bus				-.133		.203	.101	.647	.111	.124		.173		-.127
deg_doct_resch_pct_hist	.117			.196	.333			.596	-.100	-.179				
deg_bach_pct_bus	-.296	-.221		-.276	-.111	.326		.465	.169			.115	.285	
deg_doct_resch_pct_eng_libarts_vis_arts	.100	.156			.288	.149		.404	-.120	.383	-.304	-.232	-.169	
deg_doct_resch_pct_psych	-.219	-.273			.212			.403	-.239				.235	-.172
pct_Fac_FT				.104					.825					
land_grant	.159	-.110		.216			-.143	-.248	.511	-.154	.140	-.180		-.407
Pct_tot_employees_FT_fac	-.362	.102	.106	-.191	.106		-.216	-.345	-.374	.351				.107
deg_doct_resch_pct_phys_sci	-.106	.131							-.116	-.820		-.161	-.117	
deg_bach_pct_eng_libarts_vis_arts		-.176		-.184	.434				-.301	.512			-.105	.189
hosp_meddeg	.515						-.106			.213	.566	-.157	.149	
engineering_presence		.107			-.332	.242		-.112			.476			-.194
deg_doct_resch_pct_bio	.161			-.171	.383	-.222		-.438			.457			
pct_ftft_ug_rec_instl_grnt_aid			-.168	.186	-.206						.147	-.692		
frosh_admit_rate	-.286	-.321		-.242	-.137		-.250					-.606	-.175	
pct_ftft_ug_rec_st_loc_grnt_aid				-.133		.332	.287	.161	.149	.160	.188	.454		-.211
deg_doct_resch_pct_math							-.266	-.175					-.643	-.265
stu_pt_GRAD_enroll	.254	.122		-.375		.339	-.204		-.230	.298			.461	-.103
deg_doct_resch_pct_phil		.354		.114	.263			.330	-.182	.215			-.378	
Pct_SE_RD_Federal			.143		.145			-.155				-.145	.160	.773
Percent of Variance Explained	20.1	8.4	8.1	6.5	6.2	5.1	3.4	3.4	2.9	2.9	2.9	2.8	2.5	2.4
Cummulative Variance Explained	20.1	28.5	36.6	43.1	49.3	54.4	57.8	61.2	64.1	66.9	69.8	72.6	75.2	77.5

Appendix D - Dendogram



BENCHMARKING FOR STUDENT LEARNING OUTCOMES USING CASE EXAMPLES

Thomas P. Judd
United States Military Academy

Christopher Pondish
City University of New York

Charles Secolsky
County College of Morris

Of recent, benchmarking has become more important for making educational decisions. Yet, little is known about how judgments surrounding the benchmarking process reflect valid curricular decisions for student learning outcomes. One thing we realize is that benchmarks are naturally occurring phenomena that are a subset of standard-setting methods and as such may be overly used without careful consideration of the role that human judgment plays in selecting the benchmark. We believe that the use of benchmarking should be restricted to circumstances where the focus is not on politically redistributing or redefining rewards in education but only for improving student learning outcomes. To this end, this paper examines how some issues associated with the benchmarking process can be addressed.

Benchmarking is a process that can take place at many different levels at a given institution. While the most common forms of benchmarking take place among similar institutions, another type of benchmarking can be considered intra-institutional. This paper addresses benchmarking intra-institutional student learning outcomes using case examples. The findings of the study illuminate the point that when the outcomes statements associated with the mission of the institution are standards-based and not comparative then benchmarking can take place with respect to institutional standards or competencies. Another form of intra-institutional benchmarking can occur when students of different majors are compared with respect to common core skill areas in what can be referred to as normative assessment. Both types of the intra-institutional *context*s for benchmarking, standards-based or normative, depend on the mission-related institutional standards of performance. Issues identified relate to the potential for inappropriate or invalid inferences being made from outcomes assessment results using rubrics and baselines due to a lack of statistical applications.

This paper is a study of internal benchmarking of student learning outcomes using comparisons of the same or similar learning outcomes for individual courses within an institution and for the same courses over time for the purposes of formative assessment or improvement. According to Upcraft and Schuh (1996) and Seybert, Weed and Bers (2011), there are three types of benchmarking: internal, generic and comparative. Intra-institutional benchmarking is internal benchmarking.

Intra-institutional benchmarking of student learning outcomes at a given college or university is difficult to accomplish. Inter-institutional may be even more difficult to accomplish across institutions. If benchmarking is to become more valuable within and across institutions then a greater degree of standardization is necessary. For student learning outcomes, standardization can refer to the curriculum, intended student learning outcomes, evaluation and/or assessment instruments, mode of instruction (online vs. face-to-face), and testing.

Intra-institutional benchmarking involves making comparisons between units within the same institution. Earlier work on norm-referenced testing could be considered a backdrop for the development and conceptual framework of the benchmarking movement (Upcraft & Schuh, 1996). Background on intra-institutional benchmarking with respect to learning outcomes comes out of the criterion-referenced literature where each item of a test or a task is defined by some domain of interest and success on that domain over time. Optimally, the item or task being benchmarked would be representative of some set of admissible observations. In cases such as the use of rubrics and the criteria imposed by the rubrics, the variance of the scores associated with the observations can be parsed into three categories: variance attributable to criteria; variance attributable to raters; and, variance attributable to their interaction.

The Difference between a Benchmark and a Standard

For benchmarks, disaggregation of data fosters utility in that there are multiple criteria for making judgments about success. Benchmarks can more often than not be naturally occurring (e.g., comparison to best in class). A standard, on the other hand, is based on human judgments and the harnessing of those judgments to arrive at a cutscore or minimum level of acceptable performance. The fact is that a score of 65 is passing is based on judgment which has been accumulated over a number of years. Standards can be adjusted through validation by a consensus of external expertise, as determined by disciplines with national or regional accrediting agencies, or, in the case of university systems, a common core of system wide standards (Judd & Keith, 2011). A standard can also be validated by examining scores on a commonly accepted external criterion. Success based on such criteria would indicate that students have met or exceeded the standard set and thus it would mean that students are predominantly scoring successfully on the criterion for which the standard was based.

Once benchmarks have been identified, the next question is what legitimate uses can be made of intra-institutional benchmarking data on student learning outcomes? Before this question can adequately be answered, there are a number of obstacles that need to be overcome, especially if decisions made based on benchmarking data can be used to effectuate change. Foremost, faculty development initiatives are needed to overcome resistance (possibly related to issues of academic freedom, etc.) to benchmarking. Even if faculty has accepted that assessment information can document the need for improvement, making changes to the curriculum based on benchmarking is still a delicate balancing act. Resistance can also result from politically charged comparisons among the institutional departments or disciplines. Outcomes assessment has traditionally meant closing the loop after an intervention has taken place. Typically, the intervention can be theoretical and curricular validity may be unknown or lacking since student ability is often not controlled. The same can be true for differences between classes of the same course with different instructors or classes from different institutions with different learning environments

Types of Intra-Institutional Benchmarking for Student Learning Outcomes

One type of standards-based benchmarking seeks to determine how good the learning outcomes need to be (see Stake, 2004). A second type of benchmarking is represented as a criterion of performance growth over time using baselines (see Baldridge, 2011). A third type of benchmarking can take place with respect to attitudes as with indirect measures of student learning outcomes such as the National Survey on Student Engagement (NSSE). For the first type of benchmarking, the answer to how much is good enough requires that we find a point on the skill or ability continuum that represents adequate or expert attainment for the skill or ability one is assessing. For intra-institutional benchmarking of student learning outcomes, defining such a benchmark does in fact, require some form of standard-setting. The field of standard-setting in educational measurement is based on judgment and is mostly empirical. Different methodologies are used to accomplish this purpose (see Pitoniak & Morgan, 2011).

Methodology Used for Uncovering Important Issues

Three case examples were used to illuminate issues surrounding intra-institutional benchmarking. The first case is a comparison of rubric scores for a Graphic Design course and a Photography course offered by the same Graphic Design Department at a community college. The same rubric was used for both courses, which were capstone courses. The second case example shows the progress of students over a course sequence in mathematics. It exemplifies the creation of a trend and establishes a baseline by providing pass rates of students starting in a developmental Intermediate Algebra course through a course in College Algebra through Pre- Calculus through Calculus. The point of the second case is to demonstrate the advantages and disadvantages using pass rate trends as benchmarks. Finally, benchmarking is discussed from the point of view of the NSSE data on a large sample, and the potential for misleading interpretations of intra-institutional comparisons.

Table 1: - OUTCOMES ASSESSMENT RUBRIC FORM

Component Possible 100 Points	**Outstanding 25 Points**	**Highly Successful 20 Points**	**Successful 15 Points**	**Not Yet Successful 10 Points**
Technique	Very good understanding of different media and their uses. Work exhibits mastery of visual arts techniques.	Good understanding of different media and their uses. Work exhibits good control of visual arts techniques.	Solid understanding of different media and their uses is not very broad. Work exhibits competence of visual arts techniques.	Understanding of different media and their uses is not evident. Work exhibits limited mastery of visual arts techniques.
Design	Very good understanding of the elements good design and composition and uses these, skillfully and effectively to communicate ideas.	Good understanding of the elements good design and composition and uses these very well to communicate ideas in most instances.	Solid understanding of the elements good design and composition. Communication established but unintended.	Understanding of the elements good design and composition is not evident. Communication skills are poor.
Creativity and Concept	Work is unique and presents an original, interesting and clear conceptualization of an idea.	Work is mostly unique and presents a largely original, interesting and clear conceptualization of an idea.	Work contains unique and derivative elements and presents a partially original, interesting and clear conceptualization of an idea.	Work is derivative, uninteresting and lacks clarity.
Presentation	Work exhibits mastery of skills and materials without error.	Work exhibits appropriate use of skills and materials without significant errors.	Work exhibits a rough approximation of what is appropriate, includes a few errors.	Work exhibits critical errors in the use of materials or skills specific to the task.

The rubric used in Table 1 has four criteria or dimensions: Technique, Design, Creativity and Concept, and Presentation. Discretized continuous point allocations with descriptions appear in each of the 16 cells of the rubric. For both the Graphic Design and Photography courses, there were the same four judges or raters. Averages in the form of means were computed for each criterion across the judges.

For the pass rates on the same six embedded questions in the mathematics case examples, tests containing the six embedded questions were given as part of final exams in the four course sequence to students at the end of the semester and did not count in students' grades. Items were studied one at a time since each item represented a different domain of skill or ability. The problem addressed by the embedded questions was to what extent should students as an aggregate answer each of the six questions correctly? Three consecutive semesters of data: spring 2009, fall 2009, and spring 2010 were used in the math case example. Analyses produced results for each item by course and as a subtest of all six items by course.

Issues in intra-institutional benchmarking are brought forth and the difficulties encountered by comparing NSSE benchmarks across college departments are discussed. With very large numbers of cases at some institutions and the larger number of cases for the peer groups, differences using t-tests are often significant. For this purpose, effect sizes are employed with NSSE data.

Results:

Rubric scores for the Graphic Design and Photography courses are presented in Table 2 and Table 3, respectfully. The rubric scoring may be typical of rubric scoring associated with portfolio assessments for determining impact of instruction on a particular curriculum. For each criterion, as is the case for Graphic Design and Photography, limited statistical comparisons are often made. The authors surmise that this is the case because meaning is imputed in the rubric cells and an attempt is made by scorers of rubrics to keep statistical analyses simple, reflecting at most the mean or average criterion score, and at times providing data on inter-rater-reliability or agreement apart from the central analysis of the rubric scoring.

Table 2:

Student	Technique	Design	Creativity and Concept	Presentation	Total Points
1	95	90	85	100	370
2	100	90	95	100	385
3	90	90	90	100	370
4	90	85	85	100	360
5	87	95	89	95	366
6	80	85	85	95	345
7	94	100	100	100	394
8	94	95	95	100	384
9	94	95	95	95	379
10	99	100	100	95	394
11	89	90	90	95	364
12	100	95	100	100	395
13	90	95	95	95	375
14	89	85	90	95	359
15	95	100	99	100	394
16	90	85	90	95	360
17	80	90	85	95	350
18					
Average	1556/4=389 389/17=22.88 22.88x4=91.5 91%	1564/4=391 391/17=23 23x4=92 92%	1568/4=392 392/17=23 23 x4=92 92%	1658/4=413.75 413.75/17=24.33 24.33x4=97.32 97%	

Table 3:

Student	Technique (Avg.)	Design	Creativity	Presentation	Total Points (# of reviews)
1	72 (18)	70 (17.5)	71 (17.75)	70 (17.5)	283 (4) = 70.75%
2	102 (20.4)	103 (20.6)	106 (26.5)	102 (20.4)	413 (5) = 82.6%
3	91 (22.75)	89 (22.25)	82 (20.5)	83 (20.75)	345 (4) = 86.25%
4	85 (21.25)	89 (22.25)	84 (21)	85 (21.25)	343 (4) = 85.75%
5	93 (18.6)	98 (19.6)	97 (19.4)	100 (20)	388 (5) = 77.60%
6	90 (22.5)	90 (22.5)	88 (22)	96 (24)	364 (4) = 91.00%
7	87 (21.75)	84 (21)	75 (18.75)	81 (20.25)	327 (4) = 81.75%
8	53 (13.25)	63 (15.75)	63 (15.75)	65 (16.25)	244 (4) = 61.00%
9	87 (21.75)	78 (19.5)	67 (16.75)	93 (23.25)	325 (4) = 81.25%
10	76 (19)	84 (21)	86 (21.5)	78 (19.5)	324 (4) = 81.00%
11	76 (19)	72 (18)	65 (16.25)	80 (20)	293 (4) = 73.25%
12	92 (18.4)	102 (20.4)	102 (20.4)	100 (20)	396 (5) = 79.20%
13	65 (16.25)	77 (19.25)	79 (19.75)	75 (18.75)	296 (4) = 74.00%
Total Points	252.9 / 13 = 19.45	259.6 / 13 = 19.96	256.30 / 13 = 19.71	261.9 / 13 = 20.15	4341 / 55 = 79%
Average	19.45 / 25 = **77.8%**	19.96 / 25 = **79.84%**	19.71 / 25 = **78.84%**	20.15 / 25 = **80.6%**	= **79.27%**

There were 17 work products for Graphic Design and 13 work products for Photography. An inspection of the rubric scoring for Graphic Design shows mean ratings ranging 91.5 to 97.32 for the four criteria with Presentation having the highest mean. In Table 3 (to a large extent the same criteria) means ranged from 77.8 to 80.6. The same four judges rated the work assignments higher for Graphic Design on average than the student work products for Photography. But does this inference tell the whole story of the benchmarking of these two courses? Our answer is an emphatic “no.” First, are there students who have outlier performances that lower the mean score considerably? Look at student #8 for Photography – a score of 53 on Technique, or the low scores for students #8 and #9 for Creativity.

Obviously, the basic statistical concept of a standard deviation of rubric scores for each criterion would potentially provide some pivotal information if these portfolio assessments were used to assess outcomes. In fact, a generalizabilty theory G study would enable the parsing out of variance attributable to criterion effect, rater effect , and their interaction term (Webb, Shavelson & Steedle, 2011) This would help to identify instances where outcomes assessment interventions could be more effective than when the rater effect is carrying or masking the differences between criteria (see Secolsky & Judd, 2011).

Table 4

1. Solve the equation $\frac{2}{5x}+3=6$ $\boxed{\text{Answer } \frac{2}{15}}$

	M016	M110	M123	M131
a) $\frac{2}{45}$	5.8%	4.1%	1.9%	0.0%
b) $\frac{2}{15}$	47.9%	61.7%	88.7%	97.0%
c) $\frac{45}{2}$	5.8%	2.1%	0.0%	3.0%
d) $\frac{15}{2}$	21.8%	16.2%	7.5%	0.0%
e) None of the above	18.8%	15.9%	1.9%	0.0%

Table 5

2. 117 is 65% of what number? $\boxed{\text{Answer } 180}$

	M016	M110	M123	M131
Spring 2009	66.7%	75.4%	89.8%	91.7%
Fall 2009	34.4%	72.4%	88.2%	84.1%
Spring 2010	59.8%	74.2%	83.5%	88.1%

Table 6

3. $(5x-2)^2 =$	Answer $\boxed{25x^2-20x+4}$			
	M016	M110	M123	M131
a) $25x^2-4$	8.2%	4.1%	0.0%	0.0%
b) $25x^2-20x-4$	4.2%	10.0%	13.2%	12.1%
c) $25x^2+4$	6.6%	5.6%	0.0%	0.0%
d) $25x^2-20x+4$	29.7%	79.0%	86.8%	87.9%
e) None of the above	1.3%	1.3%	0.0%	0.0%

Table 7

4. Solve the equation $x^2-3x=28$	Answer 7 or -4			
	M016	M110	M123	M131
a) -7 or 4	8.1%	12.9%	1.9%	9.1%
b) 7 or -4	29.3%	73.3%	82.7%	90.0%
c) -7 or -4	3.8%	4.4%	1.9%	0.0%
d) 7 or 4	4.1%	2.1%	5.8%	0.0%
e) None of the above	4.7%	7.5%	7.7%	0.0%

Table 8

5. The perimeter P of a rectangular yard is 330 feet. The length is 75 feet more than twice the width. Find the width. $(P = 2l + 2w)$

Answer w = 30 feet

	M016	M110	M123	M131
Spring 2009	34.6%	57.0%	72.7%	78.3%
Fall 2009	20.9%	55.7%	76.9%	72.9%
Spring 2010	34.9%	58.7%	71.1%	80.6%

Table 9

2. 117 is 65% of what number? Answer 180

	M016	M110	M123	M131
Spring 2009	66.7%	75.4%	89.8%	91.7%
Fall 2009	34.4%	72.4%	88.2%	84.1%
Spring 2010	59.8%	74.2%	83.5%	88.1%

Table 10

5. The perimeter P of a rectangular yard is 330 feet. The length is 75 feet more than twice the width. Find the width. $(P = 2l + 2w)$

Answer w = 30 feet

	M016	M110	M123	M131
Spring 2009	34.6%	57.0%	72.7%	78.3%
Fall 2009	20.9%	55.7%	76.9%	72.9%
Spring 2010	34.9%	58.7%	71.1%	80.6%

Percentage of Students with Correct Answers

Semester	M016	M110	M123	M131
Spring 2009	52.6%	70.2%	87.5%	89.2%
Fall 2009	57.3%	69.7%	85.9%	87.7%
Spring 2010	50.9%	70.0%	80.5%	86.1%

The trends for most of the items from the analysis of the math course products indicated that higher percentages of students responded with the correct answer as the course became more advanced. This result is what was expected. However, for item 2, a smaller percentage of students in M131 (Calculus) answered the item correctly in comparison to those in M123 (pre-calculus). In this way, by benchmarking courses against each other, it was possible to identify where students who were learning more advanced material in mathematics demonstrated less of an ability to respond correctly than students from a less advanced course. The same was true for item 5 on the perimeter question. Otherwise, there was a clear progression of percent of students responding correctly to the items as the level of the course offering became more advanced. The identification of these two items provides an example of how benchmarking activities help make the full circle from assessment to planning. The next step in the process is to undertake the development of a plan to improve scoring on items 2 and 5, implement that plan and repeat benchmark assessment measures.

By looking at the percent of students responding to the non-correct distracters, benchmark trends like the ones in Tables 4-10 may help to identify differences in how items were conceptualized by students. For item 6, upon choosing the correct equation for the graph, the group in M016 (Intermediate Algebra) had only a 33.7% pass percentage for this item as compared to 92.3% for M123 (Pre-Calculus) and 93.9% (for Calculus), had a 7.2% responding to incorrect distracter (b).

NSSE Benchmarks

Intra-institutional benchmarking using NSSE data, while providing very valuable comparative information between departments for an institution as well as student characteristics, can at times be problematic for two reasons. First, there is no absolute standard – department means are compared to one another with respect to NSSE questions. While one department can exceed the mean for another department on a given question, Stake's (2004) point of how good should the outcomes be introduces the idea of the relative nature of benchmarking applied to a particular context. Coupled with this relative nature is the charged political comparisons that may develop as a result of intra-institutional benchmarking. Nonetheless, other types of intra-institutional benchmarking could be performed such as those between freshman and fourth year students, males and females and athletes and non-athletes on a given campus. Although, comparisons based on such antecedent characteristics such as gender carry their own wealth of political dynamite.

Table 11 and Table 12 show differences between athletes and others on NSSE benchmarks for males and females and first and fourth year students on select NSSE constructs representing Academic Challenge, Active and Collaborative Learning, Student-Faculty Interaction, Enriching Educational Experiences and Supportive Campus.. Six of the comparisons in the two tables produced statistical significance via an independent samples t-test, yet these potentially charged differences can be misleading.

Significance testing with the independent samples t-test is a standard method for comparing the means of each department, either to other departments or to the overall mean of the college. Significance testing, however, is influenced by the sample size. With larger sample sizes, statistical significance can be found with relatively small differences in means, which can make interpretations of meaningful differences challenging. The effect size statistic is independent of sample size, and can be more helpful in identifying differences that have practical significance. It is useful to note the relationship between sample size and effect size because it plays a role when interpreting results across classes, departments, institutions and systems. Effect sizes can be calculated using Cohen's d statistic, which is the difference between the means divided by the pooled standard deviations. Effect sizes between .2 and .5 are considered small, between .5 and .8 are considered medium and effect sizes of .8 or over are considered large effects. Effect sizes can be calculated for each comparison of means where statistical significance is found using the independent samples t-test.

The data in Tables 11 and 12 represent a sample of 1,734 students, qualifying it as a large sample size, as described above, with the attendant challenge of interpreting practical significance in a manner that does not distort the findings. Effect size calculations for the six statistically significant comparisons were all less the threshold of .2 for small effects, rendering the differences between the groups as relatively meaningless, and certainly not worthy of extended institutional action. They are sufficiently noteworthy to monitor over the course of several years.

Table 11
Means of Athletes* and Other Students on NSSE Benchmarks

	N	Combined	First Year	Seniors
Academic Challenge	Athlete	61.9	61.8	62.2
	Others	61.9	61.6	62.2
Active and Collaborative Learning	Athletes	54.0	51.4	59.4
	Others	54.2	51.0	57.4
Student-Faculty Interaction	Athletes	49.6	**45.3**	58.4
	Others	48.8	**41.6**	55.9
Enriching Educational Experiences	Athletes	39.3	31.2	**55.4**
	Others	41.0	30.3	**51.5**
Supportive Campus Environment	Athletes	**64.2**	64.8	**63.1**
	Others	**62.0**	64.7	**59.2**

Statistically significant differences between athletes and others means are in **bold.*** For the NSSE analysis, *athletes* are defined as those students who report yes to the item *Are you a student-athlete on a team sponsored by your institution's athletics department***.**

Table 12
Means of Male and Female Students on NSSE Benchmarks

	N	Combined	First Year	Seniors
Academic Challenge	Male	61.8	61.6	62.1
	Female	62.2	62.0	62.5
Active and Collaborative Learning	Male	**54.4**	51.3	**58.7**
	Female	**52.2**	50.3	**54.8**
Student-Faculty Interaction	Male	49.4	43.6	57.3
	Female	47.6	43.2	53.5
Enriching Educational Experiences	Male	39.9	30.4	52.8
	Female	41.5	32.3	53.8
Supportive Campus Environment	Male	63.3	64.8	61.1
	Female	62.0	64.6	58.5

Statistically significant differences between male and female means are in **bold.**

Discussion

The issue of comparing departments to each other: comparing each department to the overall mean of the institution, and giving each department only their own data. They are free to share and compare on their own, if they wish. Some important questions to be considered as this study progresses include: Is it possible to equate outcomes measures? Can the instruction, task, test, and item design be made comparable? Can item response theory be used to allow for sample-free ability estimation so that students' scores can be compared? All these hypothetical questions need to be considered as viable avenues for the future as the link between student learning outcomes and budgetary constraints take on greater importance.

Benchmarkers should pay particular attention to the use of benchmarks when dealing with nominal variables and other qualitative measures. Bearing in mind that the operationalization of multi-element, judgmental conceptual measurements is always a highly subjective if not questionable practice, the authors recommend benchmarking such measurements is, perhaps, not as viable an option as the current acceptable practices may suggest. Additional research and discussion of this point is merited.

References:

Baldridge (2011). Criteria for performance excellence 2011-2012, Gaithersberg, MD, National Institute of Standards and Technology.

Judd, T. & Keith, B. (2011) Student learning outcomes at the program and institutional levels . In C. Secolsky & D.B. Denison (Eds.) *Handbook on measurement, assessment, and evaluation in higher education.* New York: Routledge.

Pitoniak, M, & Morgan, D. (2011) Setting and validating cutscores for tests. In C. Secolsky & D.B. Denison (Eds.) *Handbook on measurement, assessment, and evaluation in higher education.* New York: Routledge.

Secolsky, C. & Judd. T. (2011). *Rubrics and outcomes assessment: A Generalizability theory approach.* Paper presented a the Annual Forum of the Association for Institutional Research, Toronto, Canada.

Seybert, J., Weed, E. & Bers, T. (2011). Benchmarking in higher education. In C. Secolsky & D.B. Denison (Eds.) *Handbook on measurement, assessment, and evaluation in higher education.* New York, Routledge.

Stake, R.E (2004) *Standards-based and responsive evaluation*. Thousand Oaks, CA. Sage.

Upcraft, M.L. & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioners*: San Francisco: Jossey-Bass Higher and Adult Education.

Webb, N., Shavelson, R. & Steedle, J. (2011). Generalizability theory in assessment contexts. In C. Secolsky & D. Brian Denison (Eds.). *Handbook on measurement, assessment, and evaluation in higher education.*

Course Withdrawal, Course Repetition, and Time to Degree:

Implications for Campus Policy

Carol VanZile-Tamsen

University at Buffalo

Paper Presented at the Annual Meeting of the

Northeastern Association of Institutional Research,

Boston, December 6, 2011.

Office of Institutional Analysis
536A Capen Hall
Buffalo, NY 14260
Email: cmv3@buffalo.edu

Abstract

Student attainment of four-year degrees is of increasing concern to educators and politicians. Past research has examined both the causes of delayed graduation, as well as interventions aimed at students to decrease time to degree. This study explores the impact of university policies related to course withdrawals and repetitions on time to degree. Suggestions for policy revision are provided.

Introduction

Prior to this study, little was known about the impact of withdrawn and repeated courses on student success (e.g., retention and graduation rates) at the University at Buffalo (UB). At UB, policies related to course withdrawal and course repetition are very lenient. After the drop-add period ends, students may withdraw from any course without an advisor's signature, although financial aid may be affected. The policy concerning repeated courses states that the grade from the first repetition of a course will be used in subsequent calculations of grade point average (GPA). If the student attempts the course a third time for an even higher grade, she may do so but only the grade from the second attempt will be counted in the GPA calculation and credits earned may or may not count toward any degree requirements. There is no limit on the number of times a student may take the same course.

This study was designed to indentify the characteristics of UB students most likely to withdraw and/or repeat courses, and to examine the impact of withdrawn and repeated courses on time to degree. It was expected that repeat registrations for the same course, either because the student was withdrawing or failing, would result in an increased amount of time to graduation and that the amount of time added could be quantified in terms of semesters. It was hoped that results would inform revisions to institutional withdrawal and repeat polices.

Importance of Degree Completion

Undergraduate degree completion is more important than ever for gainful employment among young adults. Not only do the median earnings of college graduates exceed those of people without degrees, but the fastest growing jobs in the United States in the future will require some form of post-secondary training (Hurd, 2011; Lacey & Wright, 2009). Degree completion is not only important for the employment outcomes of individual students, however. It is also important to the economic vitality of communities, states, and countries. In the United States, there has been a renewed focus on college completion rates as a result of President Obama's push to stay competitive in the global market place (de Nies, 2010). Currently, only 40.4% of US citizens ages 25 to 35 hold four-year degrees as compared to 56% of Russians, Koreans, and Canadians. Obama's goal is to increase the US percentage to 60% by 2020. The state of New York is ranked 5th among all the states in terms of college completion rates (College Board Advocacy and Policy Center, 2011), but with 48% completion, it still falls below President Obama's goal.

SUNYs new Chancellor, Nancy Zimpher, has raised the bar by focusing on the relationship between degree production and economic growth and revitalization in *The Power of SUNY*, the updated strategic plan for the State University of New York (2010). In this plan, not only is there an emphasis on graduation rates but a renewed concern with time to degree, as well as degree completion for diverse populations of students. As a result of the increased attention on college completion at both the federal and the state levels, there is an increased concern with completion rates at UB, where the 9-year average graduation rate for first-time freshmen is 62.4% with an average of 4.7 years to graduation (Academic Planning and Budget, 2011).

Jones (2011) suggests that anything that adds more time to the degree program will diminish the likelihood of graduation as "life gets in the way" (p. 4). The longer students stay enrolled in college, the stronger the competing external demands become (e.g., family life or employment). Thus, steady academic progress upon enrollment is essential to degree completion. As SUNY Chancellor Zimpher (2011) stated in a recent letter to the editor of the *Chronicle of Higher Education*: "The smoother the path to a college degree, the sooner graduates are starting businesses, filling jobs, buying homes, paying taxes, and contributing to our nation's economic recovery."

Predictors of Time to Degree

Many factors related to the students' background characteristics and achievement have been linked to college attrition and time to degree. Specifically, males, students from under-represented groups, students from low-income families, first-generation college students, and students with lower high school achievement and SAT scores are less likely to stay enrolled in college and to graduate (Adelman, 2006; Ishitani, 2006; Snyder, Dillow, & Hoffman, 2008).

Once students are enrolled, poor performance in college classes can lead to decreased momentum, and even to academic dismissal, and can ultimately increase the number of required terms (Adelman, 2006). Those who take lighter than a full-time credit load, either because they work or because they want to maintain their grades, also risk extending the degree program (Adelman, 2006; Wei, Horn, & Weko, 2009). Stopping out or transferring are also related to an increased time to degree (Wei, et al., 2009). Although Adelman suggests that it is the "swirling" from one institution to another, rather than purposeful transferring, that poses the biggest detriment to on-time degree completion.

Financial aid (or lack of it) may also play a role in time to degree. Perna (1998) found that, in the beginning of a student's college career, the type of financial aid package received was related to academic momentum. Students who received grants and work-study were more likely to make academic progress and to persist than students who received other forms of aid. This finding offers support for anecdotal accounts suggesting that students will sacrifice academic progress to avoid loans and/or to avoid taking out more loans.

Campus policy and time to degree. As Jones (2011) and Zimpher (2011) suggest, any aspects of a college experience that interfere with steady progress toward degree completion will impact time to degree. As a result, Jones suggests that colleges and universities must examine policies that have the potential to add more time. Policies regarding withdrawal from courses and repeating courses for better grades are prime suspects in the case of lengthier degree programs (Conklin, 1997). When students drop courses and/or repeat a course more than once, they are increasing the number of courses they will ultimately take to earn their degrees (Reed, 1981). While there are many personal reasons for withdrawing or repeating a course several times (Dunwoody & Frank, 1995; Ruthig, Perry, Hall, & Hladkyj, 2004), institutional policies can determine the ease and, thus, the frequency with which they do so. Specifically, at UB, there is no limit on the number of times a student can withdraw from a course and no limit on the number of times a course can be re-taken, it is likely that many students, functioning within the freedom that these policies allow, are unnecessarily extending the length of time to earn a degree.

The Present Study

Many UB faculty and staff members believe that policies related to course withdrawal and course repetition are extending degree programs beyond the traditional four years. The present study examines the extent of course withdrawal and repetition, as well as the characteristics of students most likely to withdraw or repeat courses, and the length of time added to a program due to course withdrawals and repetitions. The underlying campus "theory" is that those students who typically take longer to graduate are the same students withdrawing from and repeating courses numerous times. It is unknown if these students are also those who are more likely to drop out (i.e., males, students from under-represented groups, students from low-income families, first-generation college students, and students with lower high school achievement and SAT scores; Ishitani, 2006; Snyder, Dillow, & Hoffman, 2008). The results will be used to inform changes to campus policy.

Method

Data Source and Participants

The data used in the present analyses were taken from the University at Buffalo data warehouse for all students who entered as first-time, full-time freshmen beginning in fall, 2000, through fall, 2010. These data include variables from the following categories: demographic characteristics (i.e., gender, ethnicity, first-generation college status, family income), academic background (high school average and SAT scores), financial aid, academic progress (GPA, credit hours for each semester enrolled, degrees received), and registration history (i.e., withdrawn, failed courses, and repeated courses). The resulting sample consisted of 35,891 students.

The University at Buffalo (UB), a State University of New York University Center and a member of the American Association of Universities, currently enrolls around 19,000 undergraduates in 110 degree programs. Around 3,000 new full-time freshmen enter each year, and the university awards approximately 4,000 undergraduate degrees per year. The undergraduate population is 48% female and 59% White. Although most students (78%) come from within the state, 17% are international students. UB is largely a commuter campus with only 25% of students living in on-campus residence halls or apartments. Characteristics for the current sample are shown in Table 1.

Statistical Analyses

To address the degree to which typically at-risk students withdraw and repeat courses relative to comparison groups, univariate analyses of variance (ANOVA's) were computed for categorical variables and correlation coefficients were computed for continuous variables (Anderson & Finn, 1996; Garson, 2010). Cohen's *d* was used to calculate the effect size for each ANOVA, and variance accounted for by the relationship (r^2) was used to calculate the effect size of each correlation (Garson, 2010). These effect sizes were then used to determine the most important group differences in the number of course withdrawals, failures, and repeats.

To examine the impact of withdrawn and repeated courses on time to degree, a regression analysis (Cohen & Cohen, 1983; Tabachnik & Fidell, 2007) was conducted with the following categories of control variables: demographic background and high school achievement variables (Ishitani, 2006; Snyder, Dillow, & Hoffman, 2008); financial aid status (Cornwell, Lee, & Mustard, 2003); and academic achievement and progress toward degree at the college level (Volkwein & Lorang, 1996). The resulting standardized Beta coefficients were used to calculate the increase in semesters required with each withdrawn and repeated course.

Operational definitions of variables used. The demographic variables used in the analyses included gender (dummy coded as 1 for males and 0 for females); underrepresented minority status (dummy coded as 1 for underrepresented and 0 for not underrepresented); citizenship (dummy coded as 1 for international students and 0 for domestic students); income level (based on Expected Family Contribution after evaluation of the Free Application for Federal Student Aid or FAFSA, dummy coded as 1 for no expected family contribution); and first-generation status (coded as 1 for students whose parents had no college and 0 for all others, based on the FAFSA). The high school academic achievement variables used in the analyses included high school average, SAT Math score and SAT Verbal score. Although type of financial aid received is related to timely degree completion, it is likely that the unmet need is the factor that most influences attrition. As a result, financial aid status was operationalized as the total amount of unmet need for the first four years of study.

Variables representing academic progress at the college level included average credit hours enrolled per semester, number of changes in major program of study, and cumulative grades. The number of “Fs” received was also included. The predictors of interest were the number of withdrawn and repeated courses. Since the number of withdrawn and the number of failed courses were both extremely skewed, recoded versions (where the maximum of 6 was used to indicate 6 or more such courses) were used in the predictive model. The dependent variable, time to degree, is the number of semesters the student was enrolled from the cohort entry date to the degree date. (Students who did not file a FAFSA or who did not graduate are not included in the regression analyses).

Results

The extent of withdrawing from and repeating courses among students in the sample was surprising. Around half of the students (54.7%) had withdrawn from at least one course, and nearly all had registered for at least one course more than once (99%) while enrolled at the university[1]. On average, students withdraw from 1.31 courses (*SD* = 2.03) and repeat 10.36 courses (*SD* = 6.71). As expected, there are significant differences in the number of withdrawn courses for typically at-risk groups of students (see Table 2). Males, students from under-represented groups, first-generation students, and those from low-income families have significantly more course withdrawals than the corresponding comparison groups. Effect sizes are largest for underrepresented status (d = 0.28), income status (d = 0.22), and gender (d = 0.19). However, these are small effect sizes, the largest representing only a quarter of a standard deviation difference.

---

[1] The large percentage of course repeaters might indicate that courses with unique content carry a uniform course prefix and number, but the format of the data makes this difficult to tease out. If the courses are different, repetition will not impact time to degree for students who have not failed or withdrawn from at least one course. However, even for those students with no fails and no withdrawals, course repetition significantly predicts time to degree (β = .27), suggesting that students are repeating the same courses.

With regard to repeated courses, a slightly different pattern is found. Males and students from low-income families have significantly more repeated courses than students in the corresponding comparison groups. However, students from underrepresented backgrounds and first-generation college students have significantly *fewer* repeated courses than their corresponding comparison groups. The effect sizes are largest for gender and for underrepresented status. Males have just over a third of a standard deviation more repeated courses than females (d = 0.36), and students not from underrepresented backgrounds have about one fifth of a standard deviation more repeated courses than students who are from underrepresented backgrounds (d = -0.22). The remaining effect sizes are extremely small, even for those comparisons where there are significant differences.

Although some students may be repeating courses because they withdrew the first time, the correlation between the two variables is positive but small in magnitude (r = 0.16). The effect size (r^2) is 0.03, indicating that they share only 3% of their variance. The number of "Fs" is more strongly related to course withdrawals (r = 0.43, see Table 3) than to the number of courses that are repeated (r = 0.04). Failed courses and withdrawn courses share 18% of their variance, while failed courses and repeated courses effectively share no variance. The correlations of the remaining control variables and withdrawn and repeated courses are quite small with negligible effect sizes.

Both withdrawn and repeated courses are significantly related to time to degree (r = 0.32 and r = 0.58, respectively). In terms of effect size, course withdrawal shares 10% of its variation with time to degree, while repeated courses and time to degree share about a third of their variation (r^2 = 0.34). Even with none of the control variables taken into account, these two predictors are related to time to degree, and from these results, it appears that it is the repetition of courses that has the biggest impact on the length of degree program.

Predicting Time to Degree

Results of the regression analysis are shown in Table 4. Since the number of failed courses is correlated with both withdrawn and repeated courses and the reason for withdrawing and/or repeating courses may be due to an F grade, the interaction terms were also included in the model (Cohen & Cohen, 1983). This model (A) accounted for 60% of the variance in time to degree in semesters. Course withdrawals, repeated courses, and the two interaction terms were significant predictors of time to degree, accounting for 13% of the variance beyond the control variables. Since the interaction terms were significant, the models, without interaction terms, were run separately for students who had failed at least one course and for students who had failed no courses.

Among students who had received at least one “F” (Model B), the model accounted for 60% of the variance in time to degree, and 23% of the variance was attributable to withdrawn and repeated courses ($\beta = 0.24$ and $\beta = 0.45$, respectively). Among students with no “Fs” (Model C), only 38% of the variance in time to degree was accounted for by the model, with 12% of that attributable to withdrawn and repeated courses ($\beta = 0.21$ and $\beta = 0.33$, respectively). In both of these models, the number of repeated courses is the stronger predictor of time to degree.

Although the results suggest that these variables are more predictive of time to degree for students who have failed at least one course, the impact of withdrawn courses on the number of additional semesters required is very similar for both groups. For both groups, on average, nearly a quarter of a semester is added on for each unit increase in withdrawn courses (0.24 and 0.21 additional semesters, respectively). The impact on repeated courses is larger, adding an additional 0.45 semesters for students with at least one "F" and an additional 0.33 semesters for students with no "Fs."

It is important to note that, for all models, average credit hours per semester is the biggest predictor overall (β = -0.43, β = -0.52, and β = -0.46, respectively). These results suggest that, for each credit hour below average (Mean = 14) a student earns per semester, nearly half of a semester is added on to time to degree. The additive effects of credit hours enrolled and withdrawn and repeated courses have a very strong impact on the amount of time it takes to earn a degree.

Discussion

Findings from these analyses suggest that withdrawing from and repeating courses is a fairly common activity at the university. As expected, there are significant differences in the number of withdrawn courses for typically at-risk groups of students. Males, students from under-represented groups, first-generation students, and those from low-income families have significantly more course withdrawals than the corresponding comparison groups. From the data it is unclear why these students tend to withdraw from courses more than their corresponding comparison groups. However, this tendency to withdraw from courses before credit hours can be accumulated seems to be one reason why these students have difficulty persisting and completing their degree programs.

Because course repetition is so common an event among this student population, the pattern of differences among subgroups is quite different than for course withdrawals. The two consistent differences are for gender and income. Males and students from low-income families have significantly more repeated courses than students in the corresponding comparison groups. However, students from underrepresented backgrounds and first-generation college students have significantly *fewer* repeated courses than their corresponding comparison groups.

These findings may be related to the fact that some students are repeating courses specifically to raise their grade point averages or increase the grade earned in a particular class when it is not completely necessary because of a course failure or previous withdrawal. Perhaps students who enter the university with high expectations for college achievement are more inclined to repeat courses to maintain those expectations than are students who have lower entering expectations.

Support for this conclusion can be found in the results of the regression analyses. These results suggest that the impact of withdrawn and repeated courses varies according to whether or not students have failed courses. Course repetition has a stronger impact on time to degree for students who have failed at least one course as compared to those who have not failed any courses. The quantified impact for students with at least one F is 0.45 semesters as compare to 0.33 semesters for those students without any Fs. The impact of withdrawn courses is nearly equivalent for students who have received at least one F (0.25 additional semesters) as compared to those with no Fs (0.21 additional semesters).

This difference may be attributed to the fact that those who are repeating a course due to an "F" grade have received no credits for the failed attempt and have, thus, wasted the time spent on that attempt. Those who repeat a course after they have already completed it successfully have earned the credits from the successful attempt, which may potentially be counted as general education credits rather than just thrown away. Since credits from the early attempts may not always be "counted" in the degree program and students may choose to take the course more than two times for an even better grade, course repetition still impacts time to degree for students with no Fs, but by a smaller amount.

Although the results of this study do allow us to quantify the effect of withdrawn and repeated courses on time to degree in terms of additional semesters required, the impact for a real student at the university may be much stronger and involve more time than just part of a semester. In truth, it will always add one full semester since this university does not offer accelerated courses. In addition, the impact will vary depending on the importance of the course for the degree program and whether or not the student received an F or withdrew before they received a final grade. The best situation for on-time degree completion would be a withdrawal with no grade in a course for which there are many possible substitutes. If the course is a general education requirement or elective and can be substituted with another offering, the student may opt to substitute a different course during the next semester of enrollment. If the student registers for more credits than the typical full-time load (15), he/she may stay on track. If the student needs to remove an F from the calculation of GPA, however, he/she must re-enroll in that same course when it is next offered or when it can next be fit in.

The situation can be much different for required courses that must be repeated. For example, a student who needs to repeat a failed or withdrawn course from spring semester has several options in order to catch up before the next fall. He/she might enroll in the class, if offered, during the summer session. If the course is not offered in the summer, registration will have to wait for the following fall or spring semester, and this may add more than one semester. Alternatively, he/she may take the course at a different institution and transfer the credits. However, a course taken elsewhere may not always count and additional time may be required to acknowledge the grade and the credits, which could impact fall registration and result in a second “lost” semester for that particular course.

Additional problems may occur for students if the course they are trying to repeat is needed to stay in sequence or is a pre-requisite for admission to the major. When students get “out of sequence” or miss important application deadlines, they may have to wait an extra year to get back in sequence or to apply for admission to a major. One failed or repeated course may actually balloon into an entire year added on the length of the degree program. For this reason, it is important for those who are considering withdrawing from a course to understand the full impact of the course withdrawal on their graduation prospects. Perhaps, in the long run, a C is not all that bad if staying enrolled and earning those credits means that they will be more likely to persist and to earn the degree they are seeking. On the other hand, if the course grade is definitely going to be an F, withdrawing may be better than failing since the end result of needing to repeat the course is the same, but the F will have an immediate negative impact on the GPA, especially if the GPA is close to a 2.0 already.

Implications for Current Practice

Findings suggest that withdrawing from courses and registering for them again later significantly lengthens the amount of time required to complete degree requirements. Thus, current policy might be changed to require an advisor’s signature prior to withdrawal. During this process, the advisor might share the results of this study, specifically the length of time that may be added to the degree program when a course is resigned and repeated. In addition, advisors should explore students’ reasons for withdrawing and suggest other options, such as tutoring or changing majors. In addition, limiting the number of courses that can be resigned and repeated would discourage the number of withdrawals and course repeats that we currently see on campus.

Since course repetition has a stronger impact on time to degree for students who have already failed at least one course, providing early intervention and continuing support to students with academic deficiencies is key to ensuring that students complete courses successfully the first time. Not only should the university review the supports that are currently in place, but also train faculty to know when referral of a student to campus support services is required.

Finally, the number of credit hours earned each semester remains the strongest predictor of time to degree. For this reason, information should be provided to students early and often concerning the impact of registration patterns on the length of time required to earn a degree. Students should be encouraged to register for at least 15 hours each semester and should be cautioned to seek help early when facing a potential final grade of F.

Implications for Future Research

The focus of the present study was to quantify the impact of course withdrawals and repeats on time to degree. As a result, the course failure variable was included merely to serve as a control variable so that its effects could be teased out of those for withdrawals and repeats. However, failures also impact grades, credits earned, and changes in major. Future analyses that seek to explain as much variance in time to degree as possible should include those interaction terms and well as the three-way interaction of course failures, withdraws, and repeats.

Although most of the demographic and high school achievement variables are significant predictors of time to degree as past research suggests (Ishitani, 2006; Snyder, Dillow, & Hoffman, 2008), their impact here was small relative to the impact of average credit hours enrolled per semester, changes in major, F grades, and course withdrawals and repeats.

Adelman (2006) suggests that students who earn at least 20 credit hours in the first two semesters of enrollment have made good progress and are on track to complete the degree program. It is possible that demographic and high school achievement variables are most predictive of outcomes at the end of the first year. They may impact time to degree indirectly through college grades, credit hours earned, changed majors, etc. Future research should attempt to understand the impact of first-year success (in terms of completing a certain number of credit hours) on persistence and degree completion, as well as how a student's background impacts first-year success and indirectly influences time to degree.

Conclusions

This study has allowed the quantification of additional time required for a degree program (expressed in units of semesters) with withdrawn and/or repeat courses, both for students who have received Fs and for those who have not. As a result, there are clear implications for the revision of university policy with regard to more stringent oversight of course withdrawals and repetitions. In addition, the fact that the impact of course repetition on time to degree is so much greater for students who have failed courses suggests that early intervention efforts need to be bolstered in an attempt to prevent failure before it happens. With policy revision and improved early intervention efforts, the time it takes students to earn a degree can be shortened.

References

Academic Planning and Budget (2011). *University cumulative graduation profile, new fall freshmen, 1990-2010*. University at Buffalo, Buffalo, NY. Retrieved March 14, 2011 from http://apb.buffalo.edu/facts/enrollment/attrition/CumGradProfileFreshmen.pdf

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

Anderson, T. W., & Finn, J.D. (1996). *The new statistical analysis of data*. New York, NY: Springer.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

College Board Advocacy and Policy Center (2011). *The roadmap for educators and policy makers to help students reach college completion*. Retrieved June 14, 2011 from http://completionagenda.collegeboard.org/

Conklin, K. A. (1997). Course attrition: A 5-Yr perspective on why students drop classes. *Community College Journal of Research and Practice, 21*, 753-759.

Cornwell, C.M., Lee, K. H., & Mustard, D. B. (2003). The effects of merit-based financial aid on course enrollment, withdrawal, and completion in college. Bonn, Germany: The Institute for the Study of Labor (IZA). Retrieved June 8, 2011 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=434543

de Nies, Y. (2010, August 9). President Obama Outlines Goal to Improve College Graduation Rate in US. ABC World News. Retrieved June 14, 2011 from http://abcnews.go.com/WN/president-barack-obama-outlines-college-education-goal-university/story?id=11359759

Dunwoody, P.T., & Frank M. L. (1995). Why students withdraw from classes. *Journal of Psychology, 129*, 553- 558.

Garson, G.D. (2010). Univariate GLM, ANOVA, and ANCOVA. In *Statnotes: Topics in multivariate analysis*. Retrieved October 23, 2010 from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm .

Hurd, N. (2011, September 13). *Defining higher education success.* Retrieved October 18,2011 from http://www.usnews.com/education/best-colleges/best-practices/articles/2011/09/13/defining-higher-education-success .

Ishitani, T.T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *The Journal of Higher Education, 77*, 861-885.

Jones, S. (2011). Freedom to Fail? The board's role in reducing college dropout rates. *Trusteeship,19* (1) 2-5.

Lacey, T. A., & Wright, B. (2009). Employment outlook: 2008-18: Occupational employment projections to 2018. *Monthly Labor Review, 132* (11), 82-123.

Perna, L.W. (1998). The contribution of financial aid to undergraduate persistence. *Journal of Student Financial Aid, 28* (3), 25–40.

Reed, J. G. (1981). Dropping a college course: Factors influencing students' withdrawal decisions. *Journal of Educational Psychology, 73*, 376-385.

Ruthig, J.C., Perry, R. P., Hall, N. C., & Hladkyj, S. (2004). Optimism and attributional retraining: Longitudinal effects on academic achievement, test anxiety, and voluntary course withdrawal in college students. *Journal of Applied Social Psychology, 34*, 709-730.

Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2008). *Digest of Education Statistics 2007* (NCES 200 8022).Washington, DC: Government Printing Office.

State University of New York (2010). *The power of SUNY: Strategic plan, 2010 & beyond*. Retrieved may 1, 2011 from http://www.suny.edu/powerofsuny/pdf/SUNY_StrategicPlan.pdf.

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Pearson Education, Inc.

Volkwein, J. M., & Lorang, W.G. (1996). Characteristics of extenders: Full-time students who take light credit loads and graduate in more than four years. *Research in Higher Education, 37*, 43-68.

Wei, C. C., Horn, L., & Weko, T. (2009). *A profile of successful Pell Grant recipients: Time to degree and early graduate school enrollment (NCES 2009-156)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Zimpher, N. (2011, June). SUNY policy reforms help students transfer easily. *Chronicle of Higher Education*. Retrieved June 15, 2011 from http://chronicle.com/article/SUNY-Policy-Reforms-Help/127910/.

Table 1

Descriptive Statistics for Sample (Demographic Characteristics and Academic Achievement)

		Percentage of:					Average		
Cohort	**N**	**Female**	**Under-Represented Minority**	**International**	**Low Income**	**First-Generation**	**SAT Verbal**	**SAT Math**	**High School Average**
2000	3057	45.0	13.4	8.8	20.6	31.7	550	570	88.5
2001	2990	45.2	13.7	8.5	22.1	29.5	550	580	89.3
2002	3029	46.3	12.7	7.6	23.8	28.2	550	580	89.7
2003	3574	45.7	10.1	8.2	18.0	28.1	550	580	89.3
2004	3172	48.2	12.7	7.9	19.3	28.3	560	580	90.6
2005	3216	47.4	12.7	9.8	23.2	23.8	560	590	90.4
2006	3460	47.8	10.5	11.5	17.3	27.8	550	590	90.6
2007	3257	46.9	12.7	12.9	22.4	25.7	560	590	91.2
2008	3378	47.3	12.1	16	29.8	26.5	550	600	91.4
2009	3404	46.7	10.9	16.7	NA	24.2	550	600	91.7
2010	3354	45.7	12.6	17.7	NA	NA	550	600	92.0
Overall	**35,891**	**46.6**	**12.2**	**11.5**	**21.8**	**27.4**	**553**	**587**	**90.4**

Table 2

Univariate Subgroup Comparisons of Withdrawn and Repeated Courses

	Withdrawn Courses					Repeated Courses				
Subgroup	**Mean**	**SD**	**F**	**p**	**d**	**Mean**	**SD**	**F**	**p**	**d**
Males	1.49	2.20	330.77	0.000	0.19	11.48	7.22	118.21	0.000	0.36
Females	1.10	1.80				9.08	5.80			
Underrepresented	1.81	2.22	264.05	0.000	0.28	9.23	6.24	161.68	0.000	-0.22
Not Underrepresented	1.25	1.96				10.68	6.65			
International	1.28	2.21	0.63	0.429	-0.01	10.18	7.72	3.32	0.068	-0.03
Domestic	1.31	2.01				10.39	6.56			
Low Income	1.69	2.31	206.72	0.000	0.22	10.87	6.89	25.65	0.000	0.08
Not Low Income	1.25	1.96				10.35	6.78			
First-Generation	1.63	2.21	94.25	0.000	0.14	10.78	6.52	26.49	0.000	-0.07
Not-First Generation	1.35	2.03				11.24	6.65			
Overall	***1.31***	***2.03***				***10.36***	***6.71***			

Table 3

Correlations of Repeated and Withdrawn Courses with Control Variables and Time to degree

	Withdrawn Courses		Repeated Courses	
Variable	**r**	**r2**	**r**	**r2**
High School Average	-0.22	0.05	0.16	0.03
SAT Verbal Score	-0.03	0.00	0.06	0.00
SAT Math Score	-0.06	0.00	0.28	0.08
Total Unmet Need	0.09	0.01	0.04	0.00
Average Credit Hours Enrolled	-0.12	0.01	0.08	0.01
Cumulative College Grade Point Average	-0.25	0.06	0.21	0.04
Changes in Major	0.28	0.08	0.21	0.04
Failed Courses	0.43	0.18	0.04	0.00
Time to degree	0.32	0.10	0.58	0.34

Table 4

Multiple Regression Analyses Predicting Time to degree

	Model A		Model B		Model C	
Predictor	**β**	**p**	**β**	**p**	**β**	**p**
Gender	-0.034	0.000	-0.049	0.000	-0.044	0.000
Underrepresented Status	0.046	0.000	0.080	0.000	0.024	0.022
Citizenship	0.010	0.154	0.032	0.005	0.001	0.945
Income Status	0.029	0.000	0.037	0.002	0.039	0.000
First-Generation Status	0.015	0.029	0.008	0.459	0.024	0.017
High School Average	-0.081	0.000	-0.095	0.000	-0.084	0.000
SAT Verbal Score	0.038	0.000	0.078	0.000	0.033	0.000
SAT Math Score	-0.086	0.000	-0.094	0.000	-0.089	0.000
Total Unmet Need	-0.032	0.000	0.006	0.554	-0.076	0.000
Average Semester Hours	-0.425	0.000	-0.515	0.000	-0.457	0.000
Cumulative College Grade Point Average	0.091	0.000	0.002	0.873	0.053	0.000
Changes in Major	0.117	0.000	0.124	0.000	0.145	0.000
Failed Courses	0.110	0.000				
R^2	**0.47**		**0.37**		**0.26**	
Withdrawn Courses	0.197	0.000	0.243	0.000	0.208	0.000
Repeated Courses	0.291	0.000	0.449	0.000	0.334	0.000
Failed X Withdrawn Courses	-0.013	0.302				
Failed X Repeated Courses	0.227	0.000				
ΔR^2	**0.13**		**0.23**		**0.12**	
Total R^2	**0.60**		**0.60**		**0.38**	

DEVELOPING A CULTURE OF ASSESSMENT IN STUDENT AFFAIRS: COMPONENTS, ACTIONS AND PROCESSES

Michael N. Christakis, Ph.D.
Assistant Vice President for Student Success

Office of the Vice President for Student Success
University at Albany, SUNY

North East Association for Institutional Research (NEAIR)
38th Annual Conference

December 3 – 6, 2011
Boston, MA

**Developing a Culture of Assessment in Student Affairs:
Components, Actions and Processes**

Michael N. Christakis, Ph.D.
University at Albany, SUNY

ABSTRACT: This study identifies three theoretical frameworks which can determine if a culture of assessment exists in student affairs divisions. The frameworks are adapted for their application to student affairs assessment practices and are operationalized through a survey instrument administered to a purposive sample of student affairs assessment practitioners. The results demonstrate that leadership by the Senior Student Affairs officer (SSAO), their valuing personnel in the area of student affairs assessment, and the utilization of sophisticated, high-quality assessment practices contribute to positive culture of assessment. The results also find that a lack of resources and effective communication between staff and across organizational units can detract from a culture of assessment.

From a Movement to a Culture: From "Add-on" to "Buy-in"

Peter Ewell, in 2002, suggested that assessment in higher education was transforming from a "movement" to a "culture." As Ewell (2002) observed, early predictions indicated that "assessment would quickly go away." Whereas unsuccessful movements tend to disappear after a few years and successful ones are occasionally absorbed into more dominant cultures, for assessment, neither has occurred. While many student affairs professionals on campuses across the country are "doing something" in assessment, assessment remains an "add-on" (Ewell, 23). In most instances, student affairs professionals engage in assessment because they are instructed to do so – either by outside agents or at the urging of their college's administration.

Student affairs divisions continue to engage in assessment during professional conferences which routinely highlight or feature assessment. Increasingly, more full-time student affairs professionals are charged with administering a student affairs assessment program on their campus full-time. In a survey of student affairs professionals, nearly one quarter of respondents indicated that their student affairs division had a full-time assessment professional (Elling and Henning, 2008). Of note, however, is in those instances which a full-time student affairs assessment professional exists, they were employed at large (greater than 15,000 students), four-year, public universities; were charged with divisional assessment and were housed in a central administration (Vice President-level) office.

Any discussion of "assessment culture," however, does not begin and end with employing a full-time professional to oversee assessment efforts – divisionally or at the departmental level. While institutionalizing good assessment practices through personnel dedicated to advancing a comprehensive assessment program is ideal, given contemporary resource constraints in higher education, it may not be realistic.[1] Opting to hire a full-time assessment professional and establishing a healthy, productive assessment culture in your organization are not mutually exclusive. Indeed, organizations would be better served by both possessing personnel dedicated to assessment as well as espousing a culture of assessment that moves the organization's comprehensive assessment agenda forward. While roughly one in four student affairs divisions have a dedicated assessment professional, what remain are staff already fully engaged in their own "day to day" work to provide leadership and coordination for their offices' assessment efforts as an added responsibility.

It is in these instances – in which organizations call upon staff that are both unfamiliar and unsure of assessment work – that a true culture of assessment has the potential to move a student affairs assessment agenda forward, even without a full-time professional tasked with administering the assessment program. It is ultimately the collective values and beliefs of members of student affairs divisions, departments or offices that will have a long term impact on assessment efforts (Banta, et al., 1996).

For the better part of the last two decades, scholars and practitioners alike have sought to identify the "need" for good assessment; we have articulated the "imperative" for student learning outcomes assessment in the co-curriculum; we have encouraged the utilization of assessment findings, especially with respect to decision-making and greater transparency; and we have sought to expand our collective knowledge and expertise in the area of assessment. And yet, there is no compelling evidence that all of our collective, normative theorizing has necessarily resulted in either quality assessment or, more importantly, the emergence of a culture which might sustain good assessment practice for the next two or more decades. In fact, Elling and Henning (2008) found that fewer than 20% of student affairs assessment professionals had a good understanding of establishing a culture of assessment.

Which raises the question: what characteristics contribute to a "culture of assessment" at colleges and universities that advance a comprehensive assessment agenda? Put another way, what variables need to exist or be nurtured for a culture of assessment to take hold?

What does hospital safety have to do with it?

It seems that the sometimes arduous task of instilling a culture of assessment in student affairs, or in higher education generally is analogous to the often challenging prospect hospital

---

[1] Upcraft and Schuh (1996) identify a comprehensive model of assessment made up of the following seven elements (pgs. 27 – 30): keeping track of who uses student services, programs and facilities; the assessment of student and other clientele needs; clientele satisfaction; assessing campus environments and student cultures; assessing outcomes; comparable institutional assessment; using nationally accepted standards to assess.

administrators are confronted with when instilling a "culture of patient safety" among health care providers at their hospitals.

With that as my backdrop, I sought out scholarly writings on the issue of establishing a culture of patient safety at hospitals and found the following:

> *The safety culture of an organization is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management. Organizations with a positive safety culture are characterized by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventive measures* (Nieva & Sorra, 2003).

We could make much the same case for establishing a culture of assessment in higher education generally. To paraphrase,

> *…<u>assessment culture</u> is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of an organization's <u>assessment program</u>. Organizations with a <u>positive assessment culture</u> are characterized by communications founded on mutual trust, by shared perceptions of the importance of <u>assessment</u> and by confidence in the efficacy of <u>assessment findings</u> and their use.*

Based on our adaptation of Nieva and Sorra's (2003) insights on safety culture in hospitals, we might deduce that a "positive assessment culture" (our dependent variable) is based upon (1) communication, (2) perceptions of the relative importance of assessment to the organization, and (3) confidence in assessment findings and their use.

Similarly, Singer, et al. (2003) identify seven underlying components for establishing a culture of safety, which include (**emphasis added**):

1. **A <u>commitment</u> to safety articulated at the highest levels of the organization and translated into shared values, beliefs, and behavioral norms at all levels.**
2. **Necessary <u>resources, incentives, and rewards</u> provided by the organization to allow this commitment to occur.**
3. Safety is valued as the primary priority, even at the expense of "production" or "efficiency"; personnel are rewarded for erring on the side of safety even if they turn out to be wrong.
4. **<u>Communication</u> between workers and across organizational levels is frequent and candid.**
5. Unsafe acts are rare despite high levels of production.

6. **There is openness about errors and problems; they are reported when they occur.**
7. **Organizational learning is valued; the response to a problem focuses on improving system performance rather than on individual blame.**

For purposes of this analysis, five of Singer's (2003) aforementioned components (**bolded above**) are applied to establishing a culture of assessment in student affairs:

1. A commitment to assessment is articulated at the highest levels of the organization.
2. Necessary resources, incentives, and rewards are provided to allow this commitment to occur.
3. Communication between workers and across organizational levels is frequent and candid.
4. There is openness about assessment findings; they are shared and used in the decision-making process.
5. Organizational learning, in the area of assessment, is valued.

Building further on Singer's first component, "commitment articulated at the highest level of the organization," Singer and Tucker (2005) found that strong safety leadership required six actions:

1. Setting and communicating a clear, compelling safety vision;
2. Valuing and empowering personnel;
3. Engaging actively in the effort to improve patient safety;
4. Leading by example;
5. Focusing on system issues; and
6. Continually searching for improvement opportunities.

In the same way that we adapted Singer's (2003) components, we adapt Singer and Tucker's (2005) actions and apply them to the senior student affairs officer (SSAO) as follows:

The senior student affairs officer:

1. Sets and communicates a clear vision for assessment in student affairs;
2. Sets and communicates a compelling vision for assessment in student affairs;
3. Values personnel in the area of assessment;
4. Empowers personnel in the area of assessment;
5. Engages actively in the effort to improve programs/services based on assessment findings;
6. Leads by example;
7. Focusses on system issues by continually searching for improvement opportunities.

Returning to higher education, Suskie (2008) developed a rubric for evaluating institutional student learning assessment processes, intended as a tool to "help institutions assess the status of their current assessment efforts in terms of Middle States' accreditation standards and expectations" (2008). For the purpose of assessing the status of student affairs assessment, Suskie's (2008) criteria (dimensions) have been adapted to a more student-affairs-centered vernacular consistent with the assessment culture model introduced earlier.

The dimensions are as follows:

1. The SSAO/unit head demonstrates sustained support for promoting ongoing assessment and for efforts to improve programs and services.
2. Clear statements of expected learning outcomes at the divisional, unit and program levels have been developed.
3. Those with a vested interest in assessment are involved in developing assessments, reviewing results, and articulating findings.
4. Targets or benchmarks have been established and justified; the justifications demonstrate that the targets are appropriate given the Division's/unit's mission.
5. Multiple methods of assessment, including direct evidence, are collected and are of sufficient quality that they can be used with confidence to make appropriate decisions.
6. Evidence that has been collected is clearly linked to goals and/or learning outcomes.
7. Assessment results have been shared in useful forms and discussed with appropriate constituents, including those who can effect change.
8. Results have been used to inform planning and budgeting decisions.
9. Assessment processes have been reviewed and changes have been made to improve their effectiveness and/or efficiency, as appropriate.
10. There is sufficient engagement, momentum, and simplicity in current assessment practices to provide assurance that assessment processes will be sustained indefinitely.

Adapting Singer's (2003) components of safety culture, Singer and Tucker's (2005) actions of safety leadership and Suskie's (2008) evaluation of assessment processes, I developed an instrument which seeks to provide possible measures for determining a student affairs divisions assessment culture.

To this end, an assessment program which benefits from a positive, supportive "culture of assessment," to paraphrase Singer, et al. (2003), are characterized as possessing the following:

– ***Infrastructure****: "A commitment … at the highest levels of the organization ... resources, incentives, and rewards … to allow this commitment to occur… valued as [a] … priority…"*
– ***Outreach****: "… Communication between workers and across organizational levels is frequent and candid… Unsafe acts are rare despite high levels of production … openness about errors and problems ..."*
– ***Education****: "… Organizational learning is valued ..."*

Methodology

The survey instrument – combining the various characteristics which contribute to establishing a positive assessment culture outlined above – asked respondents their perception of assessment culture based on variations made to Singer, et al's (2003) components of establishing culture; Singer and Tucker 's (2005) leadership actions; and Suskie's (2008) evaluation of institutional assessment processes.

The instrument was administered to a purposive sample of individuals who receive electronic communications through the Student Affairs Assessment Leaders (SAAL) listerv. SAAL , established in 2008, provides opportunities for educators that coordinate assessment for divisions of student affairs to discuss issues to improve our work. SAAL's current membership includes 98 members from 34 States and two Canadian Provinces.

Twenty-nine individuals (29.5% of the sample) responded. The majority of respondents were at public four-year institutions (73%) with enrollments between 5,000 and 10,000 students and having spent, on average, between four and six years working in student affairs assessment with an earned doctorate in higher education administration.

A Lickert-like scale, with five being the highest (completely) and one being the lowest (not at all), was utilized.

Findings: Components, Actions and Processes

When analyzing the findings of the three-part instrument – including components of culture, actions of leadership and evaluation of assessment processes – ten respondents felt their division exhibited a culture of assessment either "completely" or "considerably." Of those ten institutions, nine were public four-year institutions and one was a private four-year institution. Eight had a position dedicated to student affairs assessment (six of those were considered "full-time"). Individuals charged with coordinating student affairs assessment had, on average, between eight to eleven years of experience in that area.

Of the four respondents that felt their student affairs division either only "slightly" or "did not" exhibit a culture of assessment, two were private four-year institutions, one was a public four-year institution and one was a public two-year institution. One of the institutions had a full-time position dedicated to student affairs assessment and the individual charged with coordinating student affairs assessment had worked in student affairs assessment for 10 to 12 years.

Thirteen respondents believed that their division only "moderately" exhibited a culture of assessment. For purposes of our analysis, only those respondents who felt their division possessed either a "positive" assessment culture (responding as completely or considerably) or a "negative" assessment culture (responding as slightly or not at all) will be discussed.

Singer's (2003) "Components of Culture"

For those student affairs divisions that possess a positive culture of assessment, the highest rated component is the commitment to assessment articulated at the highest levels of the organization, in these instances, by the SSAO.

Sandeen (1991) appropriately characterizes today's SSAO's as "leaders, managers, mediators, and educators." A 2008 report found that "the chief student affairs officer sets the tone for assessment in the student affairs division. This is the single most important factor determining whether the division will be successful in its assessment efforts" (Roberts and Strawn, 6). Not only does that individual need to buy-in to the need and value of assessment, they also need to make a compelling case to their staffs; staffs specialized in any variety of student services areas, but not necessarily trained or motivated to engage in assessment. [2] A transparent commitment from the SSAO for assessment is critical to establishing a culture of assessment in student affairs. "Without senior-level commitment, division staff who are resistant to assessment will have little motivation to change" (7).

Components	Positive Culture	Negative Culture
1. *A commitment to assessment is articulated at the highest levels of the organization (**INFRASTRUCTURE**).*	**4.89**	**2.75**
2. *The necessary resources are provided in order to foster a commitment to assessment (**INFRASTRUCTURE**).*	**4.22**	**2.50**
3. *The necessary resources are provided in order to foster a commitment to assessment (**INFRASTRUCTURE**).*	**3.56**	**1.50**
4. *Communication between workers and across organizational levels is frequent and candid **(OUTREACH)**.*	**3.89**	**2.25**
5. *Communication between workers and across organizational levels is candid **(OUTREACH)**.*	**4.11**	**2.50**
6. *There is openness about assessment findings; they are shared and used in the decision-making process **(OUTREACH)**.*	**4.00**	**2.50**
7. *In the area of assessment, organizational learning is valued (**EDUCATION**).*	**4.44**	**2.00**

FIGURE 1: Components of assessment culture.

Not surprisingly, even where a positive culture of assessment exists, respondents did not believe that adequate resources were provided for purposes of fostering a commitment to student

[2] *The Handbook of Student Affairs Administration* identifies departments within a division of Student Affairs as including, but not limited to, the following: Academic Advising, Admissions, Athletics, Campus Safety or Police Services, Career Development or Career Services, College/Student Union or Student Centers, Counseling Services, Dining and Food Services , Disability Support Services, Financial Aid, Fraternity and Sorority Life, Health Services, Judicial Affairs, Multicultural Affairs, Orientation and First-Year Experience, Campus Recreation, Residence Life or Housing, Spirituality, Faith or Religious Services.

affairs assessment (component **3** in **Figure 1**, above). How "resources" are defined, however, varies. Financial, human, technological and material resources all contribute to a culture of assessment.

Where a culture of assessment in student affairs exists, student affairs divisions have assessment committees in place – "human resources" – charged with advancing the assessment agenda for student affairs. While their charges may vary slightly from campus to campus, their composition is inevitably cross-divisional with representation from the various units that make-up the student affairs division. These professionals are either appointed or volunteer for membership to these committees and are charged by the SSAO with supporting good assessment practices across the division. One respondent reflected on their division's assessment committee experience, saying,

> *We have a vibrant Assessment Council in student affairs with a long history of leadership in the assessment area in Student Affairs. Most units are represented and many have more than one rep on the council. We also have reps from academic units as well since we let anyone join who wants. This has provided leadership experience for many in student affairs who do not hold positional leadership which has grown our leaders regardless of position. Student Affairs was also recognized by our accreditors with a commendation for our work in assessment. Many programs have been improved based upon our assessment data and the work that it takes to actually do meaningful assessment.*
>
> – Public, four-year; 20,000-25,000 students; Student Affairs Research/Assessment Office.

Finally, communication between staff and across organizational levels was also perceived as a weaker component in student affairs divisions with positive or negative assessment cultures (component **4** in **Figure 1**, above). Even with assessment committees bringing colleagues together, the practices exhibited by these groups warrants deeper examination. To what extent are assessment findings and practices shared across the Division? What type of progress has been made to break down long-standing silos and engage in intra-departmental assessment practices?

A respondent, commenting on the work of their division's assessment team, discussed how assessment findings were shared intra-divisionally as well as across divisions on their campus saying,

> *We…not only [have] a division wide assessment team which meets every 6-8 weeks to share assessment findings via an*

assessment 'Snapshot,' we also have formed a campus-wide assessment team where student affairs can bridge across into other units on campus.

> – Public, four-year; 30,000+ students; Student Affairs Research/Assessment Office.

Another respondent commented that staff

…are proud to see their reports being shared. Most recently we have been responding to multiple data requests from Office of Governor and our overall Board of Governors entity.

> – Public, four-year; 30,000+ students; Student Affairs Central Office (VP's Office).

In both instances above, the student affairs divisions referenced possessed positive assessment cultures.

Singer and Tucker's (2005) '"Actions of Leadership"

Of the nine actions of leadership examined, valuing and empowering personnel in the area of student affairs assessment were both rated highest in divisions of student affairs that had a positive culture of assessment (actions **5** and **6**, **Figure 2** below).

Actions	Positive Culture	Negative Culture
1. *Set a clear vision for assessment in Student Affairs **(INFRASTRUCTURE).***	**4.33**	**2.25**
2. *Communicate a clear vision for assessment in Student Affairs **(INFRASTRUCTURE).***	**4.00**	**2.50**
3. *Set a compelling vision for assessment in Student Affairs **(INFRASTRUCTURE).***	**4.11**	**2.25**
4. *Communicate a compelling vision for assessment in Student Affairs **(INFRASTRUCTURE).***	**4.22**	**2.25**
5. *Value personnel in the area of assessment **(INFRASTRUCTURE).***	**4.56**	**2.33**
6. *Empower personnel in the area of assessment **(INFRASTRUCTURE).***	**4.56**	**2.50**
7. *Engage actively in the effort to improve programs/services based on assessment findings **(INFRASTRUCTURE).***	**4.44**	**2.00**
8. *Lead by example **(INFRASTRUCTURE).***	**4.22**	**1.50**
9. *Focus on system issues by continually searching for improvement opportunities **(INFRASTRUCTURE).***	**4.33**	**2.00**

FIGURE 2: Actions of leadership.

There is no better support that the support of a peer or colleague that is confronted with similar challenges and opportunities. As Roberts and Strawn's (2008) paper states, as divisions work to "build a culture of evidence-based decision making, it is important that someone 'own' the process" (9). Never underestimate the impact of a binder and three-hole-punched handouts. In fact, the more materials personnel charged with administering assessments are provided, the more "real" their responsibilities appear. Furthermore, putting responsibilities to paper – as part of an individual's job description or "performance program" – gives the individual goals to work towards that can be evaluated and rewarded.

Increasingly, student affairs divisions have turned to in-house or third-party software platforms to help administer their assessment program. [3] Platforms empower front-line staff with the opportunity to develop and administer assessments, view data in real time, and relate findings to planning documents virtually.

For those student affairs divisions characterized as having a negative culture of assessment, the lowest rated action of leadership was the SSAO's "leading by example" (action 8, **Figure 2** above). This finding is consistent with our earlier discussion surrounding the commitment to assessment being articulated at the highest levels of the division by the SSAO. One respondent, in a division with a negative assessment culture, commented that

> *…until forced to, our [SSAO] did not even acknowledge the need for assessment let alone have anyone assigned to do assessment.*
>
> – Public, two year; 20,000 – 25,000 students; Student Affairs Research/Assessment Office.

In order to advance an assessment agenda in student affairs, the SSAO must advocate for good assessment practice. Establishing a culture of assessment begins with the organization's leader.

Suskie's (2008) "Evaluating Assessment Processes"

Similarly, when adapting Suskie's (2008) evaluation of assessment processes, the SSAO's sustained support for promoting on-going assessment as well as valuing and empowering staff (process **1** and **2**, **Figure 3** below).

Additionally, those student affairs divisions with a positive assessment culture are also characterized by their use of multiple methods of assessment of a high enough quality that they can be used with confidence to make appropriate decisions (dimensions **6** and **7**, **Figure 3** below). To advance a comprehensive assessment agenda that utilizes multiple and sophisticated methods, staff in the area of student affairs assessment need to possess the knowledge to administer such assessments.

---

[3] Since 2001, Campus Labs (formerly StudentVoice) has provided a specialized, "comprehensive assessment program that combines data collection, reporting, organization, and campus-wide integration" to over 650 colleges and universities nation-wide (www.campuslabs.com). Since 1994, EBI® (Educational Benchmarking, Inc.) has "empowered over 1,500 college and universities to impact student development, learning, retention and satisfaction through the MAP-Works® student success and retention platforms, and through national benchmarking assessments for accreditation and continuous improvement" (www.webebi.com/about).

Elling and Henning (2008) found that nearly 70% of student affairs assessment professionals had a sound grasp on "basic assessment." Furthermore, these same professionals felt comfortable in assessment design (43%), survey research (42%), and focus group administration (36%). Elling and Henning (2008) found that over 55% of student affairs assessment professionals wanted training in more sophisticated areas including data integration, data analysis, program reviews and learning outcomes.

Processes	Positive Culture	Negative Culture
1. *The SSAO/unit head demonstrates sustained support for promoting ongoing assessment and for efforts to improve programs and services* ***(INFRASTRUCTURE).***	**4.89**	**2.50**
2. *Those with a vested interest in assessment are involved in the assessment process (i.e., developing assessments, reviewing results, and articulating findings)* ***(INFRASTRUCTURE).***	**4.89**	**2.25**
3. *Targets or benchmarks have been established* ***(INFRASTRUCTURE).***	**3.22**	**1.75**
4. *Targets or benchmarks have been justified* ***(INFRASTRUCTURE).***	**3.00**	**1.75**
5. *The justifications of targets/benchmarks demonstrate that the targets are appropriate given the division's/unit's mission* ***(INFRASTRUCTURE).***	**3.00**	**1.67**
6. *Multiple methods of assessment are utilized* ***(INFRASTRUCTURE).***	**4.22**	**1.75**
7. *Assessments are of a high enough quality to be used with confidence in making appropriate decisions* ***(OUTREACH).***	**4.22**	**2.00**
8. *Evidence that has been collected is clearly linked to goals and/or learning outcomes.*	**3.89**	**2.25**
9. *There is a clear assessment plan, procedure, and/or platform for documenting the connection between evidence and goals and/or learning outcomes* ***(INFRASTRUCTURE).***	**3.89**	**1.50**
10. *Assessment results have been shared in useful forms and discussed with appropriate constituents, including those who can effect change* ***(OUTREACH).***	**4.00**	**1.75**
11. *Results have been used to inform planning decisions* ***(OUTREACH).***	**4.11**	**1.50**
12. *Results have been used to inform budgeting decisions* ***(OUTREACH).***	**3.78**	**1.25**

FIGURE 3: Evaluating assessment processes.

For the vast majority of student affairs professionals who find themselves charged with assessment as an additional responsibility, on-going training and professional development opportunities are critical to ensuring that the quality and rigor of their assessment program is maintained over the long-term. Furthermore, the organization's commitment to supporting professionals' on-going training in the area of assessment contributes to establishing a culture of assessment within their organization.

Especially given the fact that the majority of student affairs professionals engaged in assessment practice are not formally trained as assessment professionals, providing opportunities for training is tantamount to ensuring a greater understanding of assessment in student affairs. Roberts and Strawn (2008) found that “group assessment workshops are an efficient way to convey important information about the assessment process to multiple individuals, and provide an opportunity for any staff member who is interested in learning more about assessment to do so” (9). A respondent shared how their division seeks to continually hone staff’s skills in the area of assessment, suggesting they

> *… involve a wide array of SA staff in ongoing assessment capacity building by offering two workshops a month on various assessment topics. It's always surprising to see who shows up for what topics, and we've found that they're likely to follow up with us on other assessment issues once they've engaged with the workshops.*
>
> – Public, four-year; 30,000+ students; Student Affairs Central Office (VP’s Office).

Given the dwindling financial resources available to support sending professionals to conferences, campuses are faced with either committing what few dollars they have to sending one or two staff members to these gatherings or participating in distance-learning-type trainings (i.e.: webinars, teleconferences, etc.) during which more staff can participate. Student affairs divisions have begun to introduce and coordinate training opportunities on their campus. Some have gone one step further in developing certificate programs targeted at student affairs assessment which their staff can enroll and participate in throughout the year.

Those student affairs divisions that have a negative assessment culture are characterized by a relatively lower perception of clear assessment planning and documentation, as well as a lower rating of the perception of utilizing results to inform decisions (dimensions **9**, **11** and **12**, respectively, in **Figure 3** above). Divisions which espouse open and transparent planning and documentation of assessment efforts are more likely to demonstrate characteristics consistent with a positive culture of assessment. One respondent reflected that,

> *Every department is involved…on an annual basis with developing their own learning outcomes, diversity outcomes, and service and program outcomes. These assessment plans are all posted on an electronic database that any staff member can access.*
>
> – Public, four-year; 25,000 - 30,000 students; Student Affairs Central Office (VP’s Office).

Yet, assessing for assessment’s sake should never be the goal of any comprehensive assessment program. Furthermore, assessing and doing nothing with the findings does little to build confidence in the organization’s ability to analyze data as well as utilize data to make decisions about the organization’s future. Hereto, enough has not been written or studied with respect to how assessment findings are used. While there is some scholarship that examines the use of data when allocating resources, data can be used for much more than making tough financial decisions.

Two characteristics in particular are especially valuable when considering “outreach” as contributing to assessment culture. The first is “closing the loop” – that is, ensuring that those constituents that you have sought out for purposes of collecting data have some sense as to what you have done (or plan to do) with those findings. The second, “legitimacy,” is as much about the assessment program itself as it is about the data it gathers. Ultimately, both closing the loop and legitimacy are elements of greater accountability and transparency that are realized through rigorous, on-going outreach.

Beyond simply analyzing the data once it has been collected, are findings and associated decisions being shared with impacted constituencies? Further, are findings being shared with respondents? Do those students, families, staff that took the time to complete the survey, participate in the focus group or sit for an interview?

Closing the loop quite simply means sharing your findings and possibly even how those findings have helped shape decisions impacting programs and services with those individuals that contributed to the assessment. Beyond those that participated or responded as part of the assessment, broadly publicizing findings to the campus community also exposes non-respondents to findings and engenders good-will in the event they are asked to participate in a formal assessment.

What outreach also does – especially when closing the loop – is add legitimacy to an organization’s assessment activities. If respondents and related constituencies are aware that their feedback is being utilized when decisions are being made, they are more inclined to both participate in future assessments and understand the justification for certain decisions.

Conclusion

The purpose of this study was to begin to identify the variables which contribute to a culture of assessment. In this case, the subject of examination was student affairs divisions. Further to our desire to identify variables that somehow either contribute to or which help predict positive assessment culture was the challenge of identifying instruments that could measure the relative impact of these variables on assessment culture.

To that end, we identified three frameworks for evaluating culture – two grounded in patient safety and one designed to evaluate institutional assessment processes at colleges and universities. Singer, et al. (2003) identified seven underlying components for establishing a culture of safety in hospitals; Singer and Tucker (2005) found that strong safety leadership required six actions; and finally Suskie (2008) developed a rubric for evaluating institutional student learning assessment processes. Adapting Singer's (2003), Singer and Tucker's (2005) and Suskie's (2008) frameworks resulted in a new, unique instrument for determining whether a student affairs division possesses a positive assessment culture.

Our analysis revealed the following:

- **Leadership Matters**. For those student affairs divisions possessing a positive culture of assessment, the highest rated component is the commitment to assessment articulated at the highest levels of the organization, by the SSAO. For those student affairs divisions characterized as having a negative culture of assessment, the lowest rated action of leadership was the SSAO's "leading by example." Similarly, when adapting Suskie's (2008) evaluation of assessment processes, the SSAO's sustained support for promoting on-going assessment was among the highest rated dimensions for campuses demonstrating a positive culture of assessment.

- **Value and Empower Staff**. Of the nine actions of leadership examined, valuing and empowering personnel in the area of student affairs assessment, were both rated highest in divisions of student affairs that had a positive culture of assessment. In the same manner, applying Suskie's (2008) evaluation of assessment processes, valuing and empowering staff rated among the highest indicators of positive assessment culture.

- **Sophisticated, High Quality Assessment**. Those student affairs divisions with a positive assessment culture were characterized by their use of multiple methods of assessment of a high enough quality that they can be used with confidence to make appropriate decisions.

- **Still Not Enough Resources.** Even where a positive culture of assessment exists, generally respondents did not believe that adequate resources were provided for purposes of fostering a commitment to student affairs assessment.

- **The Need to Communicate**. Communication between staff and across organizational levels was also perceived as a weaker component in student affairs divisions generally, regardless of whether a culture of assessment existed. However, those student affairs divisions characterized by lower perceptions of clear assessment planning and documentation, as well as lower ratings in the utilization of results to inform decisions possessed limited or no culture of assessment.

As a preliminary study, the findings identify critical characteristics around which a culture of assessment can be built. This work only begins to scratch the surface, however, as respondents were heavily representative at large, four-year public universities. This analysis does begin to sketch out what the landscape may look like in student affairs assessment and has potential for application across any number of organizational units in higher education also engaged in assessment.

Finally, taken together the three different theoretical frameworks – reworked accordingly to apply to student affairs research and practice – provide an insightful lens through which we can examine the emergence of assessment culture in student affairs, specifically, and in higher education generally.

References

Banta, T. W., J. P. Lund and F. W. Oblander, eds. (1996) Assessment in Practice: Putting Principles to Work on College Campuses. San Francisco: Jossey-Bass.

Elling, T. and Henning G. (2008). "Student Affairs Professionals Survey: Results and Recommendations." 2008 NASPA International Assessment and Retention Conference: Scottsdale, AZ.

Ewell, P. T. (2002). "An Emerging Scholarship: A Brief History of Assessment." In T. W. Banta & Associates (Eds.), Building a Scholarship of Assessment, p. 3–25. San Francisco, CA: Jossey-Bass.

Nieva V.F., Sorra J. (2003) "Safety Culture Assessment: A tool for Improving Patient Safety in Health Care Organizations." Quality and Safety in Healthcare, 12 (Supplement 2):17-23.

Roberts, K. and Strawn, P. (2008). "Building the Infrastructure for Strong Assessment Practice in the Student Affairs Division." The Advisory Board: Custom Research Brief (April 23, 2008).

Singer S.J., Gaba D.M., Geppert J.J., Sinaiko A.D., Howard S.K., Park K.C., "The Culture of Safety in California Hospitals," Quality and Safety in Healthcare, April 2003, 12:112-8.

Singer S., Tucker A. "Creating a Culture of Safety in Hospitals." Academy of Health: Annual Research Meeting 2005. Boston, MA: June 26-28, 2005. Abstract 4197.

Suskie, L. (2008). Rubric for Evaluating Institutional Student Learning Assessment Processes. Middle States Commission on Higher Education: March 3, 2008.

Upcraft, M. L., & Schuh, J. H. (1996). Assessment in Student Affairs: A Guide for Practitioners. San Francisco: Jossey-Bass.

Measuring All Students: An Alternative Method for Retention and Completion Rates

Jason Bryer
Lisa Daniels
Excelsior College

Abstract

Retention and completion rates are frequently reported as measures of an institutions' success. However, the current method defined by IPEDS is often insufficient for many institutions, especially those who serve transfer students, continuously enroll students, or have non-traditional academic calendars. This paper outlines an alternative method that includes a larger proportion of students (e.g. all first-time enrolled students and alumni). The method relies on warehouse data so to provide a reproducible and transparent approach to calculating retention and completion rates with an emphasis on visualizations. This method also allows for more timely indications of changes in these rates. Lastly, a third measure of persistence is introduced to indicate academic activity for institutions where enrollment does not necessarily indicate academic progress.

Keywords: retention, completion, persistence, student success

Retention and completion rates are important measures of student success within an institution. However, traditional measures are often insufficient or inappropriate for institutions that have continuous enrollment and/or do not serve first-time, full-time students. This is exemplified by the definition of retention rates provided by the National Center for Educational Statistics (NCES) as part of the Integrated Postsecondary Education Data System (IPEDS). It states that retention rate is:

> A measure of the rate at which students persist in their educational program at an institution, expressed as a percentage. For four-year institutions, this is the percentage of first-time bachelors (or equivalent) degree-seeking undergraduates from the previous fall who are again enrolled in the current fall. For all other institutions this is the percentage of first-time degree/certificate-seeking students from the previous fall who either re-enrolled or successfully completed their program by the current fall.

Paper presented at the North East Association of Institutional Research Annual Conference in Boston, MA on December 6, 2011. Correspondence should be direct to Jason Bryer, 7 Columbia Circle, Albany, NY 12203-5159, jbryer@excelsior.edu.

Though for institutions such Excelsior College providing retention (and completion) rates according to the above definition is not possible, the concept of retention and completion rates are important measures that necessitate an alternative definition. This document proposes a framework for measuring retention and completion rates that maintain key features of traditional definitions but defines cohorts beyond first-time, full-time students. Specifically the features this framework maintains are:

- Define cohorts that can be followed through to 200% of normal time to completion.
- Provide a measure that is reproducible and transparent.
- Define a measure that is useful for both within and between institutions.

For institutions with continuous enrollment and/or graduation, defining cohorts provide a particular challenge. Whereas traditional institutions have very few discrete starting points (e.g. beginning of a semester), non-traditional institutions may have many starting points with overlapping semesters. It is then natural to define cohorts by a range of enrollment dates. It is important to define such ranges such that each cohort is large enough that measures of central tendency (e.g. mean, median, standard deviation) are reasonable, but also so that students at the extremes within each cohort do not have significantly different retention or completion rates. Excelsior College has approximately 966 (see figure 2) first-time enrollments[1] per month, as such defining a cohort as all students who enroll within a month is reasonable. For subpopulations where cell sizes are too small, cohorts may be combined.

Method

In order to develop an algorithm that provides reproducible results, the calculation of retention and completion rates is based upon warehouse data. Given that student information generally resides in a transactional student information system that may change as new information is obtained, warehouse data provides a "snapshot" of the student information system at a particular point in time. That is, warehouse data is merely a static copy of other data that may be transient. This method for calculating retention and completion rates relies on a warehouse table that is created once a month[2] to coincide with our cohort specification above. Specifically, the warehouse tables contain basic information about all students currently enrolled as of some particular point in time[3]. The resulting warehouse table that begins with the July 2002 cohort contains 3,427,355 records representing 102,428 unique students. Required variables in the table include: a unique student identifier (student id), student's enrollment date, and the degree the student is enrolled in as of that date. Other variables such as gender and ethnicity are only necessary if retention and completion rates broken down by these variables is desired. With these variables along with a list of graduates, the algorithm works as follows (see Figure 1):

[1]Cohorts are defined to include students enrolling in the institution for the first time. For example, students who complete more than one degree are only included in the cohort of their first degree enrollment. Moreover, students who enroll, withdraw, and re-enroll remain in their original cohort. That is, if they then complete their degree they contribute to the completion rate for that cohort.

[2]The creation of warehouse data began in February 2009. Warehouse data prior to February 2009 was created retroactively in early 2011 and therefore approximate.

[3]Excelsior College takes a "snapshot" of all currently enrolled students on the 15^{th} of each month. This has the added benefit of coinciding with IPEDS reporting date (i.e. October 15^{th}). Other reports are also

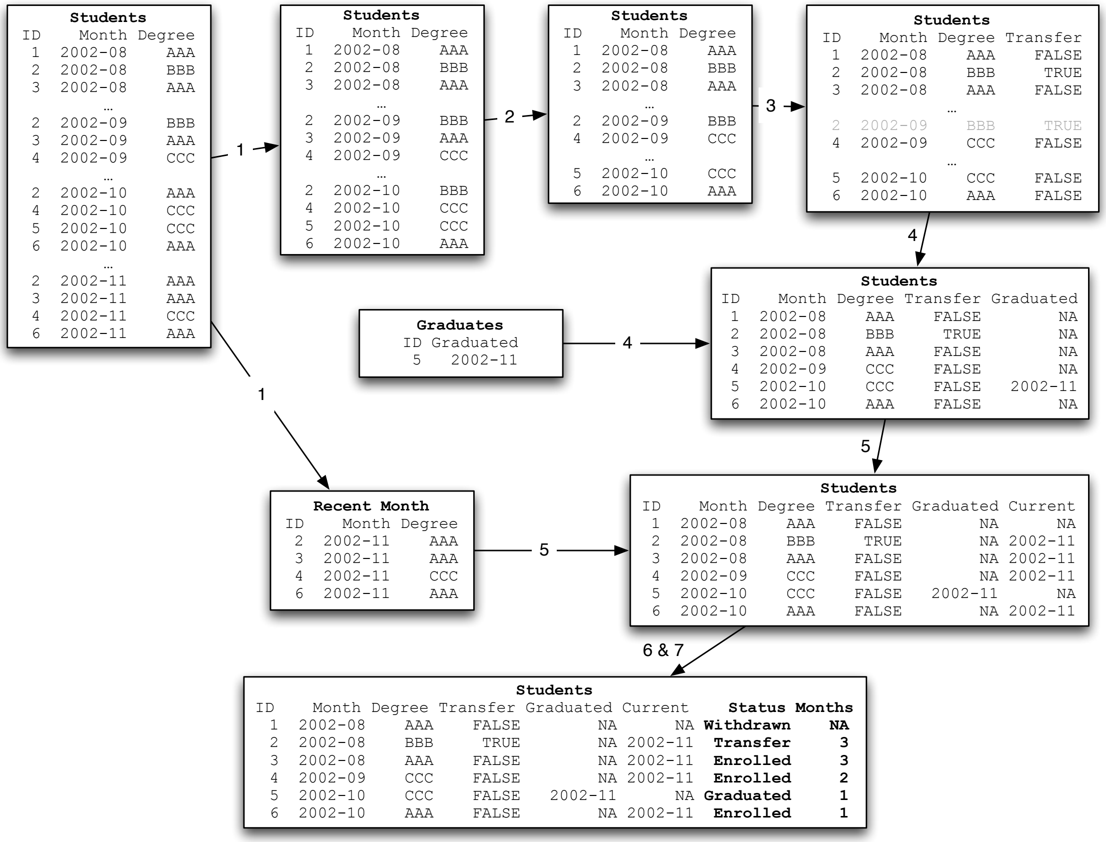


Figure 1. Graphical Representation of the Retention Algorithm

1. The most recent cohort is moved to a separate table. This will provide the basis of determining which students are still enrolled.

2. All duplicated combinations of student id and degree are removed. This leaves the earliest instance of each student and degree combination (that is, students who switched degrees will be represented more than once in the table).

3. A new variable `transferred` is calculated. For students that appear more than once in this table, the value of `transferred` is set to `true`, otherwise `false`.

4. The table is merged with the graduate table (with a `graduation date` variable) ensuring that all student records remain. The resulting table will contain a new variable with `graduation date`. If that variable is not null, then the student graduated.

5. The table is merged with the last cohort table (with a `created date` variable) created in step one above. The resulting table contains a new variable. If that variable is not null then the student is still enrolled, otherwise the student withdrew or graduated.

6. Based on the variables now available, a new factor variable (`student status`)

generated based upon this warehouse data such as persistence and demographics.

is created that classifies each student as either still enrolled, enrolled in different degree, graduated, graduated with a different degree than originally enrolled in, or withdrawn.

7. A `months enrolled` variable is calculated based upon the difference between the warehouse date (i.e. cohort) and the reference month (i.e. the most recent warehouse data).

It should be noted that the warehouse data begins in July 2002. To ensure that a student's first a occurrence in the warehouse table is their first enrollment in the college, a baseline table was created that includes all students enrolled prior to July 2002. This baseline table is concatenated to the warehouse table with a null cohort specification.

Retention and completion rates are easily calculated by aggregating the `months enrolled` and `student status` variables. A cross tabulation, divided by the cohort size, provides the rates for each cohort. Retention rates are typically reported at 15 months[4] and completion rates at 150% and 200% percent of normal time-to-completion (36 and 48 months for associate and master's degrees, and 72 and 96 months for baccalaureate degrees, respectively). These rates are calculated using weighted means[5] across all cohorts. That is, the 15-month retention is the weighted mean of each cohort's retention rate at 15 months.

Results

The overall 15-month institutional beginning retention rate as of October 2011 is 74% (this includes 89,515 students in 92 cohorts from August 2002 to March 2010). Table 1 summarizes the beginning retention and completion rates by degree level. Cell sizes are provided in parenthesis and rates are only reported in instances where the cell size is greater than 10.

		Completion Rate			
Degree Level	Retention Rate	36-Months	48-Months	72-Months	96-Months
Associate	70.89 (57210)	27.83 (46944)	32.77 (36675)	34.94 (22949)	38.20 (7418)
Bacc-Master's	54.11 (429)	1.94 (283)	5.63 (224)	4.57 (84)	3.03 (12)
Baccalaureate	80.63 (30693)	50.68 (22722)	55.60 (18376)	60.81 (10995)	65.01 (3460)
Master's	72.28 (1183)	27.35 (720)	38.68 (546)	45.23 (225)	35.66 (71)

Table 1
Aggregated Beginning Retention & Completion Rates by Degree Level

Table 2 provides a summary of cohorts from 15, 36, 48, 72, and 96 months ago. *These are the most recent cohorts that have reached the respective milestones for beginning retention and completion rates.* These rates will be much less stable then the overall, aggregated rates reported above. This is due to the smaller cell sizes as well as the variation between cohorts over time. However, these rates allow for better detection of institutional shifts in beginning retention and completion rates. Moreover, comparing cohorts over time provide some longitudinal perspective. As such, Table 2, as well as other cohort summary tables in this report, include sparklines (Tufte, 2006). Sparklines are "intense, simple, word-sized graphics" (p. 47) that provide an overall sense of the longitudinal changes in data. For our purposes here, each sparkline represents a change in rates across cohorts from older[6] on

[4]Excelsior measures retention at 15 months due to its unique enrollment model.
[5]Means are weighted by cohort size.
[6]Sparklines are limited to two years.

the left to most recent on the right. To provide context, a grey band is used to represent the overall results. Specifically, for the retention sparklines, the grey band is the overall weighted mean beginning retention rate +/- (plus and minus) two standard deviations[7]. Similarly, for the completion sparklines, the grey band is the overall weighted mean beginning completion rate at 48-months +/- (plus and minus) two standard deviations. These figures are similar to smaller versions of Figure 3 as discussed below.

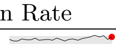
			Completion Rate				
Category	Retention Rate		36-Months	48-Months	72-Months	96-Months	Past Two Years
Associate	69.02 (510)		21.13 (426)	34.77 (604)	39.68 (499)	40.15 (675)	
Bacc-Master's	(6)		(6)	(7)	(3)	(3)	
Baccalaureate	77.85 (298)		39.10 (266)	56.69 (284)	63.24 (272)	65.74 (324)	
Master's	53.85 (13)		40.00 (20)	50.00 (12)	(7)	(8)	

Table 2
Recent Cohort Beginning Retention & Completion Rates by Degree Level

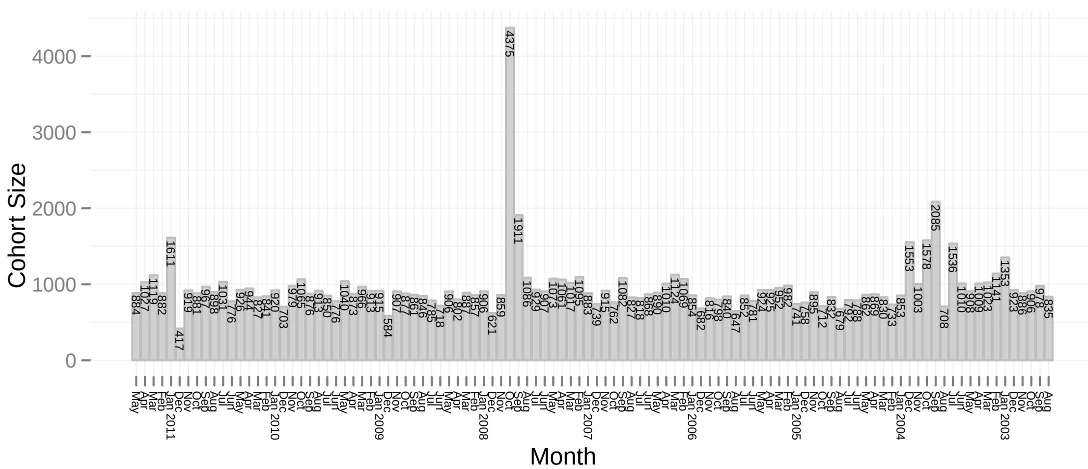


Figure 2. Cohort Sizes: First-Time Enrollments by Month

Visualizing Retention and Completion Rates

This methodology lends itself well to two approaches for visualizing retention and completion rates, namely by individual cohorts and by aggregating cohorts. Figure 3 is a cohort graphic. The x-axis corresponds to each cohort and the y-axis to percentages. That is, values on the y-axis correspond to the percentage of students within that cohort.

[7]Under normal distributions approximately 95% of values will lie within two standard deviations of the mean.

Dark-blue bars represent graduates, light-blue represents graduates but of a different degree than they first enrolled, green represents students still enrolled, light-green of students still enrolled but in a different degree than they first enrolled, and pink represents students who withdrew. Vertical lines at 15, 36, 48, 72, and 96 months highlight key milestones for retention and completion. Note that these rates correspond to a single cohort. Figure 2 is a histogram of cohort sizes and provides some context to the cohort graphic.

Figure 4 summarizes beginning retention and completion rates across cohorts. The x-axis represents months since enrollment and the y-axis, like the cohort graphic described above, corresponds to percentages. Each color represents a cohort. The solid lines correspond to beginning retention over time and the dashed lines to completion over time. The bars (dark grey for completion and light grey for retention) then represent the weighted means across all cohorts. That is, these bars provide a summary of institutional beginning retention and completion rates based upon time since enrollment. Vertical lines at 15, 36, 48, 72, and 96 months highlight key milestones for retention and completion but unlike the cohort graphic, these values are weighted means across all cohorts. Note that the instability to the right of the graph is an artifact of the decreasing number of cohorts, and therefore decreasing number of students, contributing to the average rates.

Longitudinal Rates

This methodology lends itself to multiple ways of examining longitudinal trends. Figures 8, 9, and 10 depict longitudinal trends using individual cohorts by degree level (note that the black line in Figure 8 represents the institutional beginning retention rate). However, it is typical to examine trends by fiscal year. Table 4 and Figure 6 provide 150% completion rates by degree level across fiscal years. Note that the Baccalaureate-Master's programs have been ommitted from Figures 8, 9, and 10 due to the relatively small cohort sizes and the resulting appareant volitality of rates.

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	79.80 (7337)	(13)	83.13 (3598)	70.42 (71)
2004	75.52 (9493)	86.21 (29)	83.40 (3897)	76.36 (55)
2005	75.28 (5951)	55.32 (47)	84.39 (3965)	74.49 (98)
2006	71.89 (6527)	(52)	78.48 (3787)	75.54 (139)
2007	70.30 (7223)	59.00 (100)	81.13 (3758)	76.09 (230)
2008	65.29 (10347)	45.00 (60)	79.81 (4299)	71.20 (184)
2009	63.98 (5764)	54.84 (62)	75.51 (4243)	66.21 (219)
2010	65.08 (4568)	56.06 (66)	79.88 (3146)	66.31 (187)

Table 3
Beginning Retention Rates by Fiscal Year and Degree Level

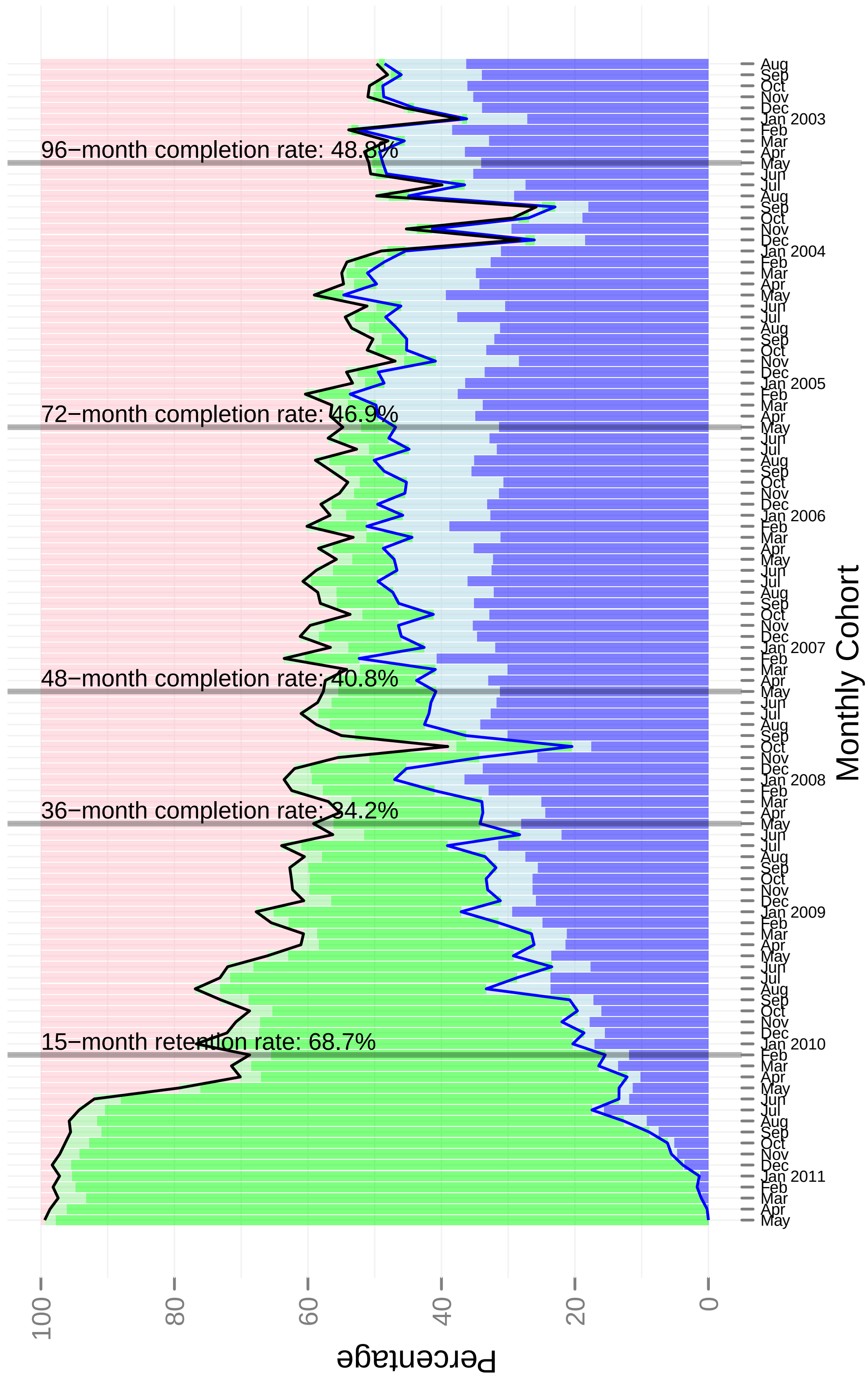


Figure 3. Institutional Beginning Retention by Cohort

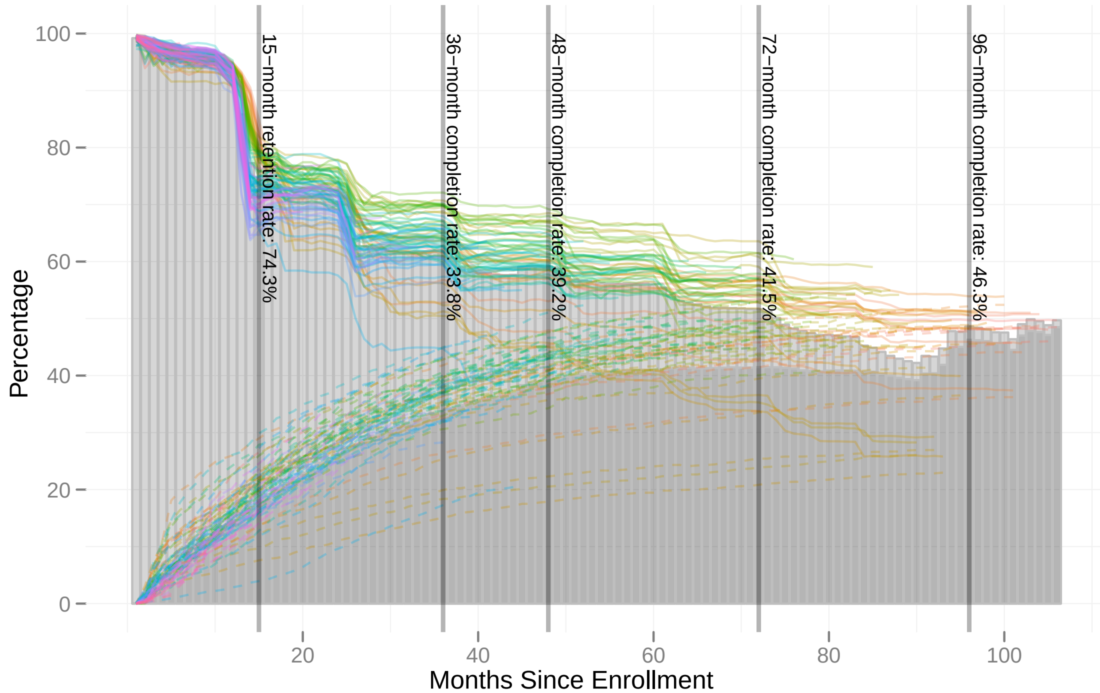


Figure 4. Institutional Beginning Retention

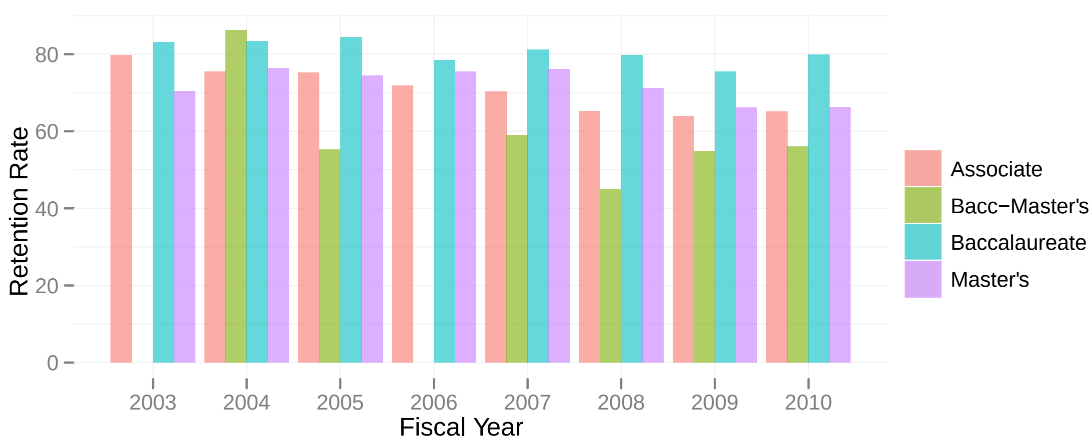


Figure 5. Beginning Retention Rates by Fiscal Year & Degree Level

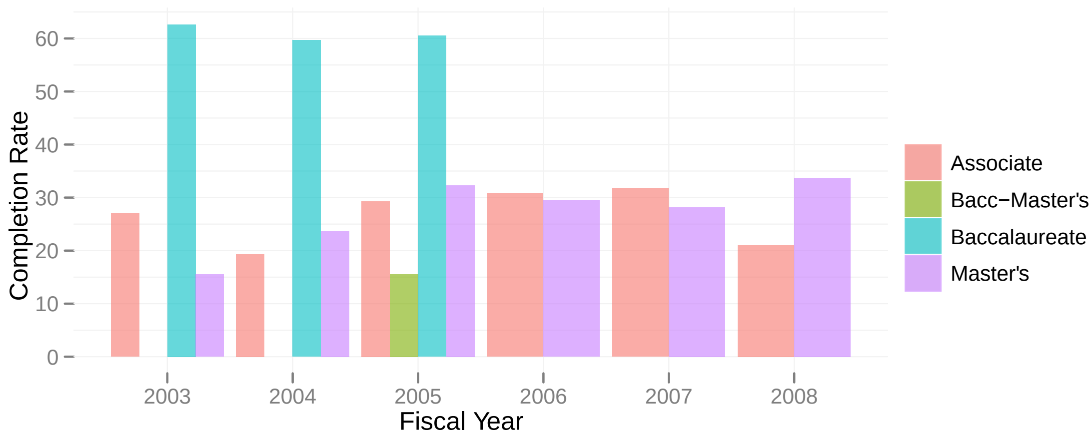


Figure 6. 150% Beginning Completion Rates by Fiscal Year & Degree Level

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	27.12 (7339)	(12)	62.64 (3472)	15.49 (71)
2004	19.31 (9498)	(27)	59.66 (3748)	23.64 (55)
2005	29.32 (5983)	15.56 (45)	60.53 (3775)	32.32 (99)
2006	30.87 (6556)			29.55 (132)
2007	31.82 (7222)			28.19 (188)
2008	21.04 (10346)			33.71 (175)

Table 4

Beginning Completion Rates by Fiscal Year (Three-year rate for Associate and Master's, Six-year rate for Baccalaureate and Bacc-Master's)

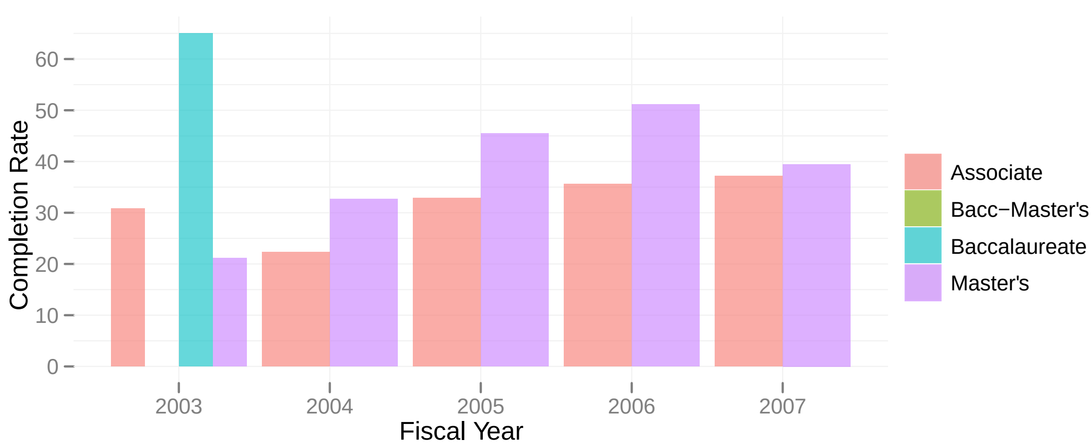


Figure 7. 200% Beginning Completion Rates by Fiscal Year & Degree Level

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	30.86 (7353)	(12)	65.00 (3460)	21.13 (71)
2004	22.36 (9527)			32.73 (55)
2005	32.90 (5978)			45.45 (99)
2006	35.63 (6561)			51.15 (131)
2007	37.16 (7256)			39.47 (190)

Table 5

Beginning Completion Rates by Fiscal Year (Four-year rate for Associate and Master's, Eight-year rate for Baccalaureate and Bacc-Master's)

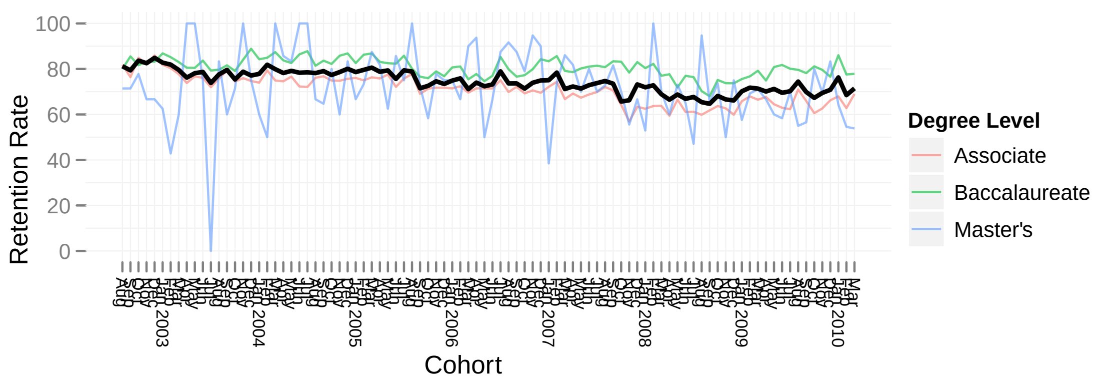


Figure 8. Beginning Retention Rates Across Cohorts

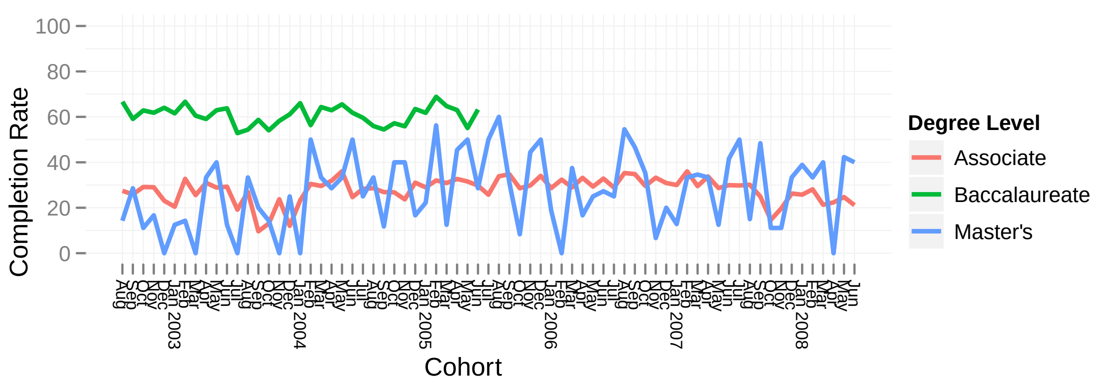


Figure 9. 150% Beginning Completion Rates Across Cohorts

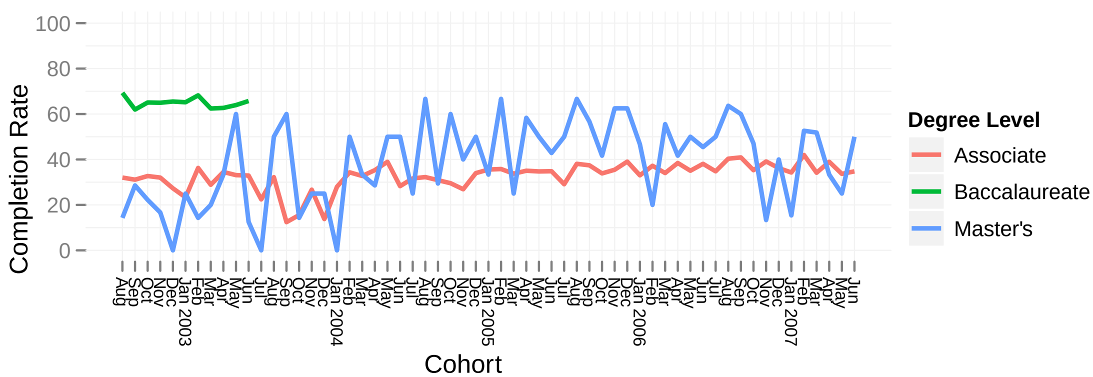


Figure 10. 200% Beginning Completion Rates Across Cohorts

Academically Active Students

Beginning retention and completion rates are important measures of student activity within an institution, however for institutions where enrollment does not necessarily imply academic progress towards a degree, these measures may be insufficient on their own. Persistence rates are a common approach to measuring the relative academic activity of students. At Excelsior College, an enrolled student is considered persistent if they have engaged in at least one of the following activities within the past six months:

- Attempted an Excelsior College course
- Attempted an Excelsior College exam (this includes portfolio assessment and military students)
- Transferred in credit
- Participated in an online conference or CPNE Workshop (Nursing)
- Participated in a preceptorship (select portion of Nursing)
- Registered for the CPNE (within 9 months)
- Completed the CCS100 Excelsior College Student Experience

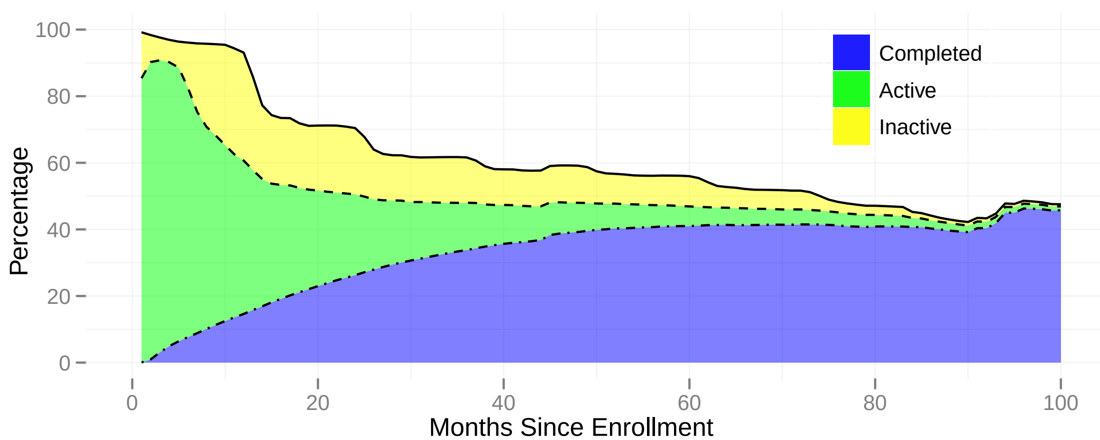


Figure 11. Beginning Completion, Retention, & Academically Active Rates

Persistence rates are calculated based upon all enrolled students at a particular time using the same warehouse data as described above. What is missing with this approach is how student persistence changes over the course of their enrollment. Figure 11 is a modified version of Figure 4 that separates retained students based upon whether they are academically active or not. Specifically, the blue region corresponds to students who have completed their degree, the solid black line corresponds to the retention rate, the shaded green area corresponds to students who are still enrolled and academically active, and the shaded yellow area corresponds to students who are enrolled but not academically active. Figure 12 provides an alternate view of these data. The solid line corresponds the percentage of students still enrolled, the green line is percentage of students who are academically active of those still enrolled, the light gray line also represents academically active students but is based upon the full cohort size, and lastly, the current months institutional persistence rate is represented by the horizontal line to provide context.

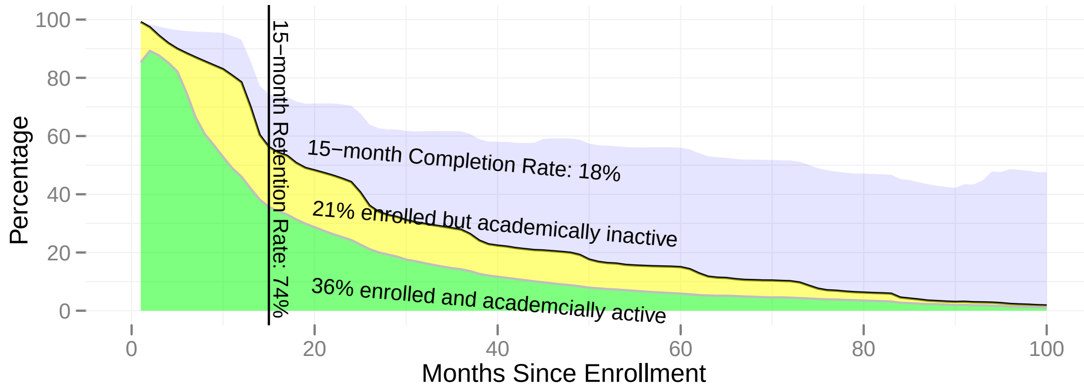


Figure 12. Academically Active Rates

Retention & Completion Rates for Alumni

The primary focus of this analysis has been on first-time to Excelsior College students. However, alumni are an important population to consider. Of the 142084 students who have graduated from Excelsior College, 7392 have returned since July 2002 to attempt a second degree. In total, 7663 students have earned more than one degree. Table 6 provides 36-month completion rates for first-time to Excelsior College students compared with alumni by fiscal year.

FY	First-Time	Alumni
2003	35.47 (10937)	(357)
2004	27.95 (13396)	(437)
2005	36.67 (10000)	(490)
2006	38.02 (10423)	(575)
2007	37.78 (11086)	(712)
2008	29.93 (14827)	(815)

Table 6
36-Month Completion Rates for First-Time & Alumni Students by Fiscal Year

Conclusions

Retention and completion are important measures for an institution to consider. However, traditional measures potentially exclude substantial portions of an institutions' student population, especially for non-traditional institutions with continuous enrollment and/or enroll students with transfer credit. The method outlined here provides an approach that includes virtually all students who attend an institution while retaining the original spirit of

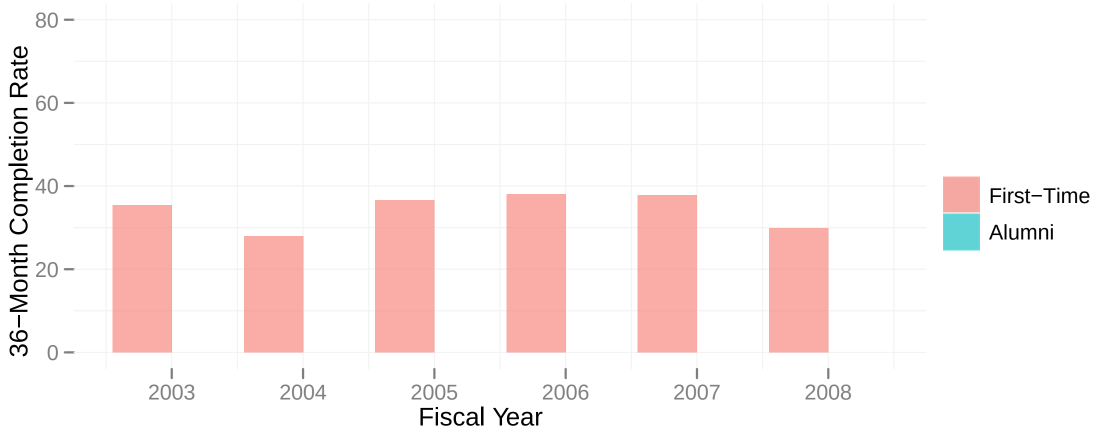


Figure 13. 36-Month Completion Rates for First-Time & Alumni Students

the traditional measures. Moreover, through the use of modern graphic techniques further insights can be achieved beyond a single numeric metric.

References

R Development Core Team. (2011). *R: a language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from R Foundation for Statistical Computing: `http://www.R-project.org/`

Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.

Wickham, H. (2011). *Devtools: tools to make developing r code easier*. R package version 0.4. Retrieved from `http://CRAN.R-project.org/package=devtools`

Appendix
Implemented for Reproducible Research

As indicated at the outset, a major goal for the method outlined here is to provide transparency and reproducibility. The algorithms outlined in this paper are implemented as a package for the open source statistical program R (R Development Core Team, 2011) and are available at `http://github.com/jbryer/irutils`. The latest development version of the `irutils` package can be installed from Github using the `devtools` (Wickham, 2011) package:

```
> library(devtools)
> install_github('irutils', 'jbryer')
```

Once installed, the `irutils` package can be loaded with the `library` command. There are two functions that perform the retention and completing rate calculations. The `cohortRetention` function returns the rates for each cohort based upon the most recent cohort whereas the `retention` function returns the rates aggregated across all cohorts.

```
> library(irutils)
> str(irutils::cohortRetention)

function (students, grads, gradColumn = "START_DATE", grouping = NULL)

> str(irutils::retention)

function (students, grads, ...)

> ecCohortRetention = cohortRetention(students, graduates)
> retention = retention(students, graduates)
```

The structure of the returned data frames are provided below.

```
> str(ecCohortRetention)

'data.frame':        106 obs. of  10 variables:
 $ Cohort          : Factor w/ 106 levels "2002-08","2002-09",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Graduated       : num  36.3 33.9 36.1 35.2 33.9 ...
 $ Graduated Other: num  12.2 12.1 12.7 13.4 10.2 ...
 $ Still Enrolled : num  0.838 1.534 1.104 1.58 0.975 ...
 $ Transferred    : num  0.359 0.511 0.883 0.79 0.542 ...
 $ Withdrawn      : num  50.3 51.9 49.2 49 54.4 ...
 $ GraduationRate : num  48.5 46 48.8 48.6 44.1 ...
 $ RetentionRate  : num  49.7 48.1 50.8 51 45.6 ...
 $ PersistenceRate: num  80 70 61.1 61.9 64.3 ...
 $ Enrollments    : int  835 978 906 886 923 1353 1141 1023 1009 908 ...

> str(ecRetention)
```

```
'data.frame':        5671 obs. of  11 variables:
 $ Cohort          : Factor w/ 106 levels "2002-08","2002-09",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ GraduationRate : num  48.5 46 48.8 48.6 44.1 ...
 $ RetentionRate  : num  49.7 48.1 50.8 51 45.6 ...
 $ PersistenceRate: num  80 70 61.1 61.9 64.3 ...
 $ Enrollments    : int  835 978 906 886 923 1353 1141 1023 1009 908 ...
 $ Graduated      : num  36.3 33.9 36.1 35.2 33.9 ...
 $ Graduated Other: num  12.2 12.1 12.7 13.4 10.2 ...
 $ Still Enrolled : num  0.838 1.534 1.104 1.58 0.975 ...
 $ Transferred    : num  0.359 0.511 0.883 0.79 0.542 ...
 $ Month          : num  106 105 104 103 102 101 100 99 98 97 ...
 $ Comparison     : chr  "2011-06-15" "2011-06-15" "2011-06-15" "2011-06-15" ...
```

Running Head: POST-BACCALAUREATE OUTCOMES

Predictive Modeling of Achievement of Post-Baccalaureate Outcomes

Tom McGuinness

North East Association for Institutional Research (NEAIR) Annual Meeting

Boston, Massachusetts

December 5, 2010

Introduction

The effects of the recent economic downturn have permeated society and its impact on postsecondary education in the United States is no exception. As state support for higher education and institutional endowment income have dwindled, the pressure on families to make up the difference in lost revenue through tuition has risen. The scenario for graduating students is no less bleak as these challenges are compounded by a tight labor market, in which students are hard pressed to find jobs after graduation. In response to this difficult climate, it is understandable the students and their families might focus college-going decisions and goals for the undergraduate experience on outcomes that will improve students' likelihood of succeeding in the labor force upon graduation. Viewing a college education as a way to improve one's career options and earning potential is not a new phenomenon. According to the Higher Education Research Institute (HERI), students in 2006 indicated that, on average, the two most important reasons for attending college were "to learn about things that interest me" and "to get a better job," the same leading reasons that students indicated in 1976 (Pryor et al, 2007). However, over this same period, the percentage of students who responded that "to be able to make more money" was a very important reason for going to college rose from 49.9 percent to 69 percent. Also, in 2006, approximately two-thirds of students indicated that "the chief benefit of a college education is that it increases one's earning power." During a time of relative economic prosperity in this country, HERI reported that approximately three-quarters of college freshmen in 1997 identified getting a better job and making more money as the most important reasons for attending college. Even as an advocate for the liberal arts, it is difficult to blame a student for not appreciating knowledge for its own end when faced with potentially hundreds of thousands of dollars of student loans and a difficult job market on the horizon. This study aims to identify the factors that increase the likelihood that a student will succeed in fulfilling his or her post-baccalaureate plans. It examines two separate post-baccalaureate outcomes, attainment of employment and admission to graduate school, for students completing their bachelor's degrees from Tufts University, a private, highly selective research university in Medford, Massachusetts. The first model examines the factors that predict whether students who pursue professional positions have accepted a position by the time they graduate. The second model considers the factors that predict students' success in graduate and professional school admissions. Additionally, I examine the graduate and professional school admission outcome further in a third model, in which the outcome is whether a graduating student is accepted to his or her first choice graduate or professional school. In particular, the study examines the relationship between student experiences, namely participation in specific co-curricular activities (independent research, research with faculty, volunteer work, study abroad) and internships, and the likelihood that students will achieve their intended post-baccalaureate plans of either gaining employment or being accepted to graduate or professional school.

Literature

The conceptual framework predominantly used to guide research on the labor market outcomes of a bachelor's degree is derived from the economic theory of human capital (Becker, 1975; Mincer, 1993). According to this theory, a student's college education is considered an investment in human capital as he or she invests time and/or money in order to develop his or her knowledge and productive skills. This theory assumes that students base decisions on their expectations about the monetary costs (opportunity cost, tuition, etc.) and benefits (higher income) of a college education. Human capital theory has guided a broad, well established body of literature on the monetary returns to postsecondary education (Paulsen, 1998). According to

Leslie and Brinkman (1988), the annual rate of return for a bachelor's degree is about 12 or 13%, a strong investment relative to alternative opportunities. Pascarella and Terenzini (2005) estimate that a bachelor's degree is equivalent to a 34 percentage point advantage in occupational status or prestige and a 20 to 40 percent advantage in earnings. Research using earnings as an outcome is fairly extensive (see Pascarella and Terenzini, 2005, for a thorough review), using a variety of predictors (e.g. institution type, grades, co-curricular engagement, work experience) and subgroup analyses (e.g. race, gender, academic field).

Though success in finding a job is frequently of great importance to students and families, the body of research on initial employability, on the other hand, is considerably smaller. In general, those who have earned a bachelor's degree are more likely to participate in the labor force and less likely to be unemployed (Paulsen, 1998; U.S. Department of Education, 1992). According to the Bureau of Labor Statistics, in 2009, the unemployment rate for those with a bachelor's degree or more was less than half of the unemployment rate of United States citizens with a lower level of educational attainment (4.6%, compared to 9.6%).

While the general labor market benefits of a college education are widely recognized, what are less clear are the within-school characteristics and factors that contribute to a student's success in finding employment upon graduation. Sagen, Dallam, and Laverty (1997) explored students' success in obtaining "baccalaureate level" employment within two months following graduation. Using a survey of more than one thousand graduates of the University of Iowa, the authors found that, compared with liberal arts majors, students in specialized majors were significantly more likely to have earned a job within two months following graduation. Additionally, the study demonstrated that work experience related to career goals was a positive, significant predictor of student success, while participation in student organizations was positively associated and marginally significant. Examining interactions between career preparation variables and personal and academic characteristics, the authors found positive effects related to participation in internships, work experiences related to career goals, advanced skills courses in quantitative analysis and writing, and participation in student organizations.

The body of research suggests that timing of measuring whether a graduate has attained a job is important. This is most apparent in examining the relationship between grades and job attainment. When using job attainment shortly after graduation as an outcome, the research is mixed, suggesting that there may be a significant positive relationship between college grades and attaining a job. In the aforementioned study by Sagen et al (1997), grade point average was a significant positive predictor of having a job within two months of graduation. However, Grayson (1997), in a study of graduates of York University in Ontario, found that there was no significant relationship between grade point average and finding a job within three months of graduation, with or without controls. If a relationship does exist, research indicates that any significance seems to disappear over time. Stoeker and Pascarella (1991), in a study of female college graduates, found that there was no significant relationship between college grades and likelihood of being employed in a full-time position nine years after entering college. Bowen and Bok (1998) reported similar results for women almost two decades after college entrance.

There are several studies that examine job attainment outcomes based on subgroup. As mentioned above, Stoeker and Pascarella (1991) observed the career attainment of women. Bowen and Bok (1998) and Perna (2005) investigated labor market outcomes based on gender and race. Also researching race-based labor outcomes, Strayhorn (2008) used data from the Baccalaureate and Beyond Longitudinal Study to examine the labor market outcomes of African American college graduates, finding that graduating from an historically Black college or university was negatively associated with earnings but positively associated with career status attainment. Also using Baccalaureate and Beyond, Bellas (2001) examined differences in job seeking patterns and outcomes based on age. One key difference with this study is that the author used number of interviews and number of job offers as outcomes. Examining the differences based on age, Bellas found that older graduates, on average, had fewer interviews and job offers, though the difference in job interviews disappeared once controls were introduced. The relationship between age and job offers persisted after the controls were added, though older graduates appeared to have been employed in better jobs when they were surveyed a year after graduation. Additionally, the analysis revealed that there is evidence that, controlling for a variety of factors, more technical majors (e.g., business management, engineering) receive the most job interviews.

The literature on the factors associated with graduate school admission appears to be even more limited. It also follows the human capital framework in that it is an investment of time and money to develop additional knowledge and skills. Zhang (2005) notes that much of the inquiry in this area of research presumes that the most important outcomes of a college education are economic, partially due to the availability of employment and earnings data in national longitudinal databases. However, this focus ignores other important outcomes such as education beyond the baccalaureate degree. To address this gap, Zhang investigated the relationship between quality of undergraduate institution and graduate education. Using Baccalaureate and Beyond data, he found that institutional quality does matter. Students from medium and high quality undergraduate institutions are more likely than those from low-quality institutions to enroll in graduate programs. Considering individual characteristics, some key findings included that females and undergraduate business majors were less likely to continue to a graduate program and there was a significant positive relationship between graduate enrollment and undergraduate grade point average. Hearn (1987) examined the post-baccalaureate educational aspirations of students at two institutions in the 1970s. In addition to pre-college characteristics, he found that grade point average, parental supportiveness, faculty-student interaction, and academic major all had a significant relationship with educational aspirations beyond the baccalaureate degree. Perna (2004) investigated gender and race differences in the decision of college graduates to pursue additional education and found that, controlling for a variety of factors, females are more likely to enroll in master's and sub-master's programs and males are more likely to enroll in first professional programs. Regarding race, taking into account expected costs and benefits, financial and academic resources, and cultural and social capital, Black graduates were more likely to enroll in a master's or professional degree program compared to Whites.

These existing studies establish the factors that are associated with whether a college graduate enrolls in a graduate or professional program. Where there is a gap in the research, though, is regarding the success of those intending to go to graduate school in achieving this goal. The current study examines students who applied to graduate school and determines the factors that predict whether they will be accepted. While such analyses are undoubtedly conducted at the institutional level, these institutional models and case studies are rarely extended into the scholarly literature.

Data

This study employs data compiled from the Tufts University Senior Survey. Administered annually to all graduating students, the Senior Survey is a comprehensive instrument that gathers a vast amount of information about the Tufts undergraduate experience.

The topics it addresses include, but are not limited to, experience and satisfaction with student services and programs, skills and abilities developed, perceptions of campus life, major and minor satisfaction, post-baccalaureate plans, community service and civic engagement, and alumni activities. The survey continues to grow and gather additional pieces of information from graduating students. In 2010, the size of the survey has swelled to more than 150 questions. Despite the scale of this survey, response rates have remained quite high. The survey regularly garners response rates greater than 90%, reaching more than 95% in each of the last 3 years. It is administered to students beginning in mid-April prior to their expected graduation in May. The survey typically remains open for about a month, closing at the time of graduation.

This study uses Senior Survey data from 2006 through 2009. The total dataset includes responses from 5,269 graduating seniors over this period[1]. A little more than half of this group is female (53.9%) and about two-thirds reported that they are White (67.6%). Students represent a broad range of majors, 75 in total, so these majors have been condensed into five categories: Arts and Humanities, Engineering, Interdisciplinary, Natural Science, and Social Science. Since students could declare more than one major, the frequencies of these variables sum to greater than 100 percent of the analytic sample. Additionally, because the majors are not mutually exclusive, each of the major category variables is included in the regression models (rather than excluding one as a comparison group). The analysis also uses data from the survey on engagement in student activities. The variables included in the analysis are independent research, research with a professor, study abroad, and participation in volunteer work. The

[1] The original dataset included a total of 5,913 responses. Grade point outliers, with values of zero (N = 51) and values greater than four (N = 2), were removed from the analysis. Additionally, forty-four cases with a missing gender were removed, as well as eight responses from transgender students. Most substantially, approximately 9 percent of cases were dropped from the analysis due to missing SAT scores. While this was a considerable number of cases to remove, it did not have any impact on the direction or significance of the coefficients in each model. As a result, I determined that it was more important conceptually to include a baseline measure of academic achievement than to have the dropped cases in the models.

dataset also includes information about student experiences with internships, including the number of internships a student has held. Also included in the dataset are variables specifying the type of jobs students intended to have after graduation and, for those who applied to graduate or professional programs, the fields of study of the programs to which they applied. The job types have been recoded from the 59 options included on the 2009 Senior Survey to eleven categories based on Bureau of Labor Statistics definitions. Similarly, the graduate fields of study have been recoded into eight categories determined by the author. Finally, the study employs cumulative undergraduate grade point average as a measure of academic achievement and cumulative SAT score (Math & Verbal) as a baseline academic achievement measure.

Since there are multiple outcomes, the analysis is split into two non-mutually exclusive groups. The first group is all students who indicated that they plan to work in the fall after graduation and had started their job search prior to graduation. The second group is all students who applied to graduate or professional school for admission following graduation. These groups are not mutually exclusive because a student could have applied for admission to graduate or professional school and then made the decision to pursue a professional position (whether he or she was admitted or not). About 60 percent of respondents indicated that they planned on employment following graduation (N = 3,147). About one-fifth of respondents reported that they had applied to graduate or professional school (N = 1,062)[2].

Tables 1 through 4 summarize the descriptive statistics of the respondent population. Table 1 presents descriptive statistics for the overall population and by post-baccalaureate plans. Table 2 displays descriptive statistics for students planning on employment following graduation,

[2] Note that the sum of these numbers is less than the 5,269 students in the dataset. Some graduating students are not included in any of the regression models because they were either undecided about their post-baccalaureate plans or, if they were planning to work following graduation, they had not yet started their job search. The descriptive characteristics of this population are included in Table 1.

broken out by whether the student had accepted a job or not. Tables 3 and 4 present descriptive statistics for students who had applied to graduate school, with Table 3 based on whether the student had been accepted to any graduate or professional school and Table 4 based on whether he or she had been accepted to his or her first choice institution.

Table 1. Descriptive Statistics, Overall and by Post-Baccalaureate Plans (2005-2009)

	Total (N = 5,269)	Employment (N = 3,147)	Grad/Prof (N = 1,062)	Neither (N = 1,184)
Student Characteristics				
Male	46.1%	46.9%	44.5%	46.0%
Female ©	53.9%	53.1%	55.5%	54.0%
Asian	12.9%	12.4%	14.7%	12.4%
Black	6.7%	7.2%	5.1%	7.4%
Hispanic	6.8%	6.4%	5.5%	8.8%
White ©	67.6%	68.7%	69.6%	62.8%
Foreign/International	3.3%	2.9%	3.3%	4.6%
Other	2.7%	2.4%	1.9%	4.1%
Major Category				
Arts and Humanities	27.8%	26.4%	25.6%	32.8%
Engineering	14.9%	15.3%	22.2%	8.2%
Interdisciplinary	10.9%	10.3%	11.8%	11.7%
Natural Science	15.6%	12.9%	26.7%	13.0%
Social Science	52.7%	56.2%	41.8%	52.7%
Student Activities				
Independent Research	39.6%	38.7%	50.7%	33.3%
Research with faculty	32.9%	32.1%	42.8%	26.9%
Study abroad	48.5%	50.7%	40.1%	48.8%
Volunteer work	48.3%	48.6%	54.0%	42.4%
0 internships ©	33.8%	27.0%	36.8%	48.5%
1 internship	26.1%	27.0%	26.3%	23.6%
2 internships	22.0%	24.5%	21.7%	15.7%
3 internships	11.8%	13.6%	11.0%	8.8%
4 or more internships	6.2%	7.9%	4.2%	3.4%
Career Plans				
Architecture and Engineering	9.0%	9.9%	12.5%	4.1%
Arts, Design, Entertainment, Sports, and Media	8.9%	8.6%	3.8%	13.9%
Business and Financial Operations	18.7%	25.7%	4.4%	12.4%
Community and Social Services	5.2%	6.0%	2.6%	5.2%
Computer and Mathematical	2.2%	3.0%	1.2%	1.2%
Education, Training, and Library	7.9%	8.6%	9.6%	5.3%
Healthcare Practitioners and Technical	5.7%	3.8%	12.9%	4.1%
Legal	4.5%	3.9%	9.5%	1.7%
Life/Natural Sciences	7.7%	7.5%	9.5%	7.1%
Social Sciences	8.0%	9.0%	6.4%	6.9%
Other/Undecided ©	22.0%	14.1%	27.5%	37.9%
Intended Field of Graduate/Professional Study				
Arts & Humanities	2.4%	0.8%	7.4%	2.2%
Engineering (including Architecture)	3.7%	1.4%	15.3%	0.3%
Law School	3.7%	1.1%	14.7%	0.9%
Medical School	3.4%	1.0%	12.6%	0.9%
Natural Science	2.3%	0.9%	7.8%	1.4%
Other Health	2.7%	1.0%	10.3%	1.4%
Social Science	2.9%	1.5%	8.9%	1.4%
Other ©	78.8%	92.5%	23.1%	91.4%
Post-Baccalaureate Outcomes				
Accepted a job	21.2%	35.5%	3.4%	0.0%
Accepted to graduate/professional school	17.7%	2.2%	87.6%	0.0%
Accepted to 1st choice grad/prof school	11.7%	1.1%	57.9%	0.0%
Academic Achievement Indicators				
Cumulative Undergraduate GPA	3.38	3.37	3.50	3.30
Cumulative SAT Score (Math & Verbal)	1,350	1,349	1,367	1,339

© = Comparison Group

Table 2. Descriptive Statistics, Seniors Planning on Employment following Graduation (2005-2009)

	Accepted position (N = 1,116)	Looking for job (N = 2,029)
Male	52.7%	43.7%
Female ©	47.3%	56.3%
Asian	13.3%	12.0%
Black	5.8%	8.0%
Hispanic	5.4%	7.0%
White ©	70.1%	67.9%
Foreign/International	2.5%	3.1%
Other	3.0%	2.1%
Arts and Humanities	19.6%	30.2%
Engineering	19.9%	12.8%
Interdisciplinary	9.3%	10.8%
Natural Science	14.9%	11.9%
Social Science	56.3%	56.2%
Independent Research	41.0%	37.4%
Research with faculty	37.3%	29.3%
Study abroad	47.7%	52.3%
Volunteer work	47.0%	49.5%
0 internships ©	24.0%	28.7%
1 internship	25.5%	27.7%
2 internships	26.9%	23.2%
3 internships	15.5%	12.5%
4 or more internships	8.1%	7.9%
Career Plans		
Architecture and Engineering	12.3%	8.5%
Arts, Design, Entertainment, Sports, and Media	4.7%	10.8%
Business and Financial Operations	29.0%	23.9%
Community and Social Services	2.9%	7.7%
Computer and Mathematical	4.8%	2.0%
Education, Training, and Library	10.9%	7.3%
Healthcare Practitioners and Technical	3.6%	3.9%
Legal	4.7%	3.5%
Life/Natural Sciences	6.5%	8.0%
Social Sciences	7.3%	10.0%
Other/Undecided ©	13.5%	14.4%
Post-Baccalaureate Outcomes		
Accepted a job	100.0%	0.0%
Accepted to graduate/professional school	2.5%	2.1%
Accepted to 1st choice grad/prof school	1.3%	1.1%
Academic Achievement Indicators		
Cumulative Undergraduate GPA	3.42	3.34
Cumulative SAT Score (Math & Verbal)	1355	1346

© = Comparison Group

Table 3. Descriptive Statistics, Seniors Planning on Graduate/Professional School following Graduation (2005-2009)

	Accepted to grad school (N = 930)	Not accepted (N = 132)
Male	43.0%	55.3%
Female ©	57.0%	44.7%
Asian	14.8%	13.6%
Black	4.2%	11.4%
Hispanic	5.6%	4.6%
White ©	70.7%	62.1%
Foreign/International	2.9%	6.1%
Other	1.8%	2.3%
Arts and Humanities	26.6%	22.0%
Engineering	22.6%	19.7%
Interdisciplinary	12.5%	6.8%
Natural Science	25.6%	34.1%
Social Science	41.9%	40.9%
Independent Research	51.3%	46.2%
Research with faculty	43.6%	37.9%
Study abroad	42.4%	24.2%
Volunteer work	55.1%	46.2%
0 internships ©	35.3%	47.7%
1 internship	26.8%	22.7%
2 internships	22.4%	16.7%
3 internships	11.3%	9.1%
4 or more internships	4.3%	3.8%
Intended Field of Graduate/Professional Study		
Arts & Humanities	7.9%	4.6%
Engineering (including Architecture)	16.3%	7.6%
Law School	16.1%	4.6%
Medical School	12.7%	12.1%
Natural Science	7.7%	8.3%
Other Health	11.2%	3.8%
Social Science	9.5%	4.6%
Other ©	18.6%	54.6%
Post-Baccalaureate Outcomes		
Accepted a job	3.0%	6.1%
Accepted to graduate/professional school	100.0%	0.0%
Accepted to 1st choice grad/prof school	66.1%	0.0%
Academic Achievement Indicators		
Cumulative Undergraduate GPA	3.53	3.27
Cumulative SAT Score (Math & Verbal)	1369	1355

© = Comparison Group

Table 4. Descriptive Statistics, Seniors Planning on Graduate/Professional School following Graduation, Accepted to 1st Choice or Not (2005-2009)

	Accepted to 1st Choice (N = 615)	Not accepted to 1st Choice (N = 447)
Male	41.1%	49.2%
Female ©	58.9%	50.8%
Asian	13.3%	16.6%
Black	3.7%	6.9%
Hispanic	5.0%	6.0%
White ©	72.4%	65.8%
Foreign/International	3.3%	3.4%
Other	2.3%	1.3%
Arts and Humanities	26.5%	25.3%
Engineering	23.3%	20.8%
Interdisciplinary	13.2%	9.8%
Natural Science	25.2%	28.6%
Social Science	40.0%	44.3%
Independent Research	51.1%	50.1%
Research with faculty	44.9%	40.0%
Study abroad	43.1%	36.0%
Volunteer work	55.6%	51.7%
0 internships ©	36.8%	36.9%
1 internship	27.3%	24.8%
2 internships	20.8%	22.8%
3 internships	10.6%	11.6%
4 or more internships	4.5%	3.8%
Intended Field of Graduate/Professional Study		
Arts & Humanities	9.8%	4.3%
Engineering (including Architecture)	18.5%	10.7%
Law School	10.1%	21.0%
Medical School	11.2%	14.5%
Natural Science	7.8%	7.8%
Other Health	13.8%	5.4%
Social Science	9.4%	8.1%
Other ©	19.4%	28.2%
Post-Baccalaureate Outcomes		
Accepted a job	2.3%	4.9%
Accepted to graduate/professional school	100.0%	70.5%
Accepted to 1st choice grad/prof school	100.0%	0.0%
Academic Achievement Indicators		
Cumulative Undergraduate GPA	3.54	3.44
Cumulative SAT Score (Math & Verbal)	1348	1371

© = Comparison Group

Methodology

This study employs binary logistic regression modeling as its primary statistical technique. This is an appropriate method for predicting a dichotomous outcome[3]. As mentioned above, the study examines three different outcomes and employs three separate models, one related to job attainment and two related to graduate and professional school admissions. Related to job attainment, the first model predicts the dichotomous outcome of whether students who plan on employment following graduation and had commenced their job search at the time of the survey had accepted a position at the time of graduation or not. Related to graduate and professional school admissions, the second model predicts the dichotomous outcome of whether students who applied to graduate school were accepted to any graduate school and the third model predicts whether these students were accepted to their first choice graduate or professional school. The job attainment model also adds specificity to this outcome by including type of job desired as a control. Similarly, the first choice graduate or professional school model includes graduate or professional field of study. The other graduate school model, acceptance to any graduate school, does not include this control due to cell size issues. For example, of the 54 students intending to attend graduate school in education, only one did not get into any graduate school. Each model employs gender and race as demographic control variables. To account for academic characteristics, major category (Arts and Humanities, Engineering, Interdisciplinary, Natural Science, Social Science) and cumulative grade point average are included as predictor variables. The key predictor variables are related to participation in academic and professional co-curricular activities: independent research, research with faculty, study abroad, volunteer activities, and internships. Regression tables are included in the body of the paper to summarize

---

[3] Note that, since the job attainment and graduate school admission outcomes are not mutually exclusive, a multinomial model is not an appropriate approach.

the results of the models. However, predicted probabilities of each outcome are likely to serve as a more practical tool for career services and academic services staff and were distributed for internal use.

Results

Job Attainment

In the analysis of the job attainment outcome, there were several significant predictors of whether a student has accepted a job at the time they responded to the Senior Survey. Regarding demographic background variables, males were about 50 percent more likely than females to have accepted position. Specifically, holding all other variables constant, males had greater odds of having accepted a job by a factor of 1.421. Controlling for other factors, there was one significant relationship between race and whether a student had accepted a job. Students with a race categorized as "Other" were significantly more likely than White students to have accepted a position (odds ratio = 1.674).

When considering the relationship between academic variables and job attainment, both academic major and cumulative grade point average were factors. Students with majors in the Arts and Humanities were significantly less likely to have accepted a position compared to those who did not have a major in these fields. Engineering majors, on the other hand, were significantly more likely to have attained a job at the time of the survey relative to those who were not Engineering majors. Controlling for other factors, they were about 60 percent more likely to have accepted a job. Students with majors in the Natural Sciences were also more likely to have accepted a position, though this predictor was only marginally significant ($p < 0.10$). Cumulative grade point average also had a significant positive relationship with the job attainment outcome variable. According to the model, holding all other variables constant, a one-point increase in grade point average increased the odds of having accepted a job by a factor of 2.762.

Considering participation in co-curricular activities, having participated in more than one internship significantly increased the likelihood of having accepted a position by graduation. Though positive, the relationship between having a single internship and having accepted a job was not significant. Having two internships or more, however, was a significant positive predictor of whether a student had accepted a job by graduation. Participation in other co-curricular activities was not a strong predictor of whether a student had accepted a job, controlling for other factors. Only research with faculty was a marginally significant predictor of job attainment. Considering a student's intended type of position, there were several variables that significantly predicted a student's success in having accepted a position. Students intending to work in Computer and Mathematical occupations and Education, Training, and Library occupations were significantly more likely to have accepted a position, while those pursuing Arts, Design, Entertainment, Sports, and Media occupations, Community and Social Services occupations, and Social Sciences occupations were less likely to have accepted a position.

Table 5. Logit Regression of Job Acceptance Among Students who Started their Job Search Prior to Graduation

	Coef.	Std. Error	p	Odds Ratio	Delta P	Sig.
Student Characteristics						
Male	0.351	0.122	0.000	1.421	0.080	***
Asian	0.041	0.126	0.733	1.042	0.009	
Black	-0.021	0.176	0.906	0.979	-0.005	
Hispanic	-0.120	0.156	0.493	0.887	-0.028	
Foreign/International	-0.289	0.188	0.250	0.749	-0.066	
Other	0.515	0.414	0.037	1.674	0.117	*
Major Category						
Arts and Humanities	-0.349	0.087	0.005	0.706	-0.080	**
Engineering	0.469	0.300	0.012	1.599	0.107	*
Interdisciplinary	0.028	0.153	0.854	1.028	0.006	
Natural Science	0.301	0.221	0.066	1.351	0.069	~
Social Science	0.134	0.153	0.318	1.143	0.031	
Student Activities						
Independent Research	0.032	0.094	0.721	1.033	0.007	
Research with faculty	0.177	0.112	0.059	1.194	0.041	~
Study abroad	-0.139	0.076	0.113	0.870	-0.032	
Volunteer work	-0.075	0.076	0.365	0.928	-0.017	
1 internship	0.079	0.119	0.474	1.082	0.018	
2 internships	0.333	0.157	0.003	1.396	0.076	**
3 internships	0.545	0.232	0.000	1.725	0.124	***
4 or more internships	0.416	0.249	0.011	1.516	0.095	*
Career Plans						
Architecture and Engineering	0.028	0.211	0.890	1.029	0.007	
Arts, Design, Entertainment, Sports, and Media	-0.754	0.090	0.000	0.470	-0.171	***
Business and Financial Operations	0.176	0.156	0.178	1.193	0.040	
Community and Social Services	-0.940	0.089	0.000	0.391	-0.212	***
Computer and Mathematical	0.601	0.455	0.016	1.823	0.137	*
Education, Training, and Library	0.581	0.296	0.000	1.787	0.132	***
Healthcare Practitioners and Technical	-0.103	0.210	0.658	0.902	-0.024	
Legal	0.394	0.322	0.070	1.483	0.090	~
Life/Natural Sciences	-0.300	0.143	0.120	0.741	-0.069	
Social Sciences	-0.363	0.121	0.036	0.696	-0.083	*
Academic Achievement Indicators						
Cumulative Undergraduate GPA	1.016	0.377	0.000	2.762	0.229	***
Cumulative SAT Score (Math & Verbal)	-0.001	0.000	0.118	0.999	0.000	

Number of Observations = 3,147

Pseudo R-square = 0.0717

Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ~ $p < 0.10$

Comparison groups: Female, White, Zero Internships, Other/Undecided Career Plans

Model also includes fixed effects by year

Acceptance to Any Graduate/Professional School

When it comes to getting into graduate school, getting good grades is critical. Based on the model, a one-point increase in grade point average increases the likelihood of being accepted to graduate or professional school more than tenfold. Academic major was unrelated to graduate and professional school acceptance, controlling for other factors. None of the demographic background variables, neither gender nor race, was a significant predictor of the graduate and professional school acceptance outcome.

Study abroad was the only co-curricular activity that was a factor in graduate admissions. Controlling for all other variables, participation in study abroad doubled the likelihood of acceptance (odds ratio = 2.003). Neither of the research variables (participation in independent research and participation research with faculty) nor volunteer engagement were significantly associated with acceptance to graduate school, holding other variables constant. Internships also were not significantly related to admission to graduate and professional school.

Table 6. Logit Regression of Graduate/Professional School Acceptance Among Students who Applied to Graduate/Professional School

	Coef.	Std. Error	p	Odds Ratio	Delta P	Sig.
Student Characteristics						
Male	-0.241	0.170	0.267	0.786	-0.026	
Asian	0.215	0.378	0.481	1.240	0.023	
Black	-0.343	0.306	0.426	0.710	-0.037	
Hispanic	0.576	0.891	0.250	1.779	0.063	
Foreign/International	-0.361	0.359	0.484	0.697	-0.039	
Other	-0.235	0.561	0.741	0.791	-0.026	
Major Category						
Arts and Humanities	-0.109	0.275	0.722	0.897	-0.012	
Engineering	0.308	0.487	0.389	1.361	0.034	
Interdisciplinary	0.477	0.647	0.235	1.611	0.052	
Natural Science	-0.407	0.215	0.208	0.665	-0.044	
Social Science	-0.154	0.262	0.615	0.858	-0.017	
Student Activities						
Independent Research	-0.156	0.203	0.511	0.855	-0.017	
Research with faculty	0.084	0.276	0.741	1.088	0.009	
Study abroad	0.695	0.483	0.004	2.003	0.076	**
Volunteer work	0.266	0.277	0.210	1.305	0.029	
1 internship	0.240	0.333	0.361	1.271	0.026	
2 internships	0.439	0.459	0.139	1.550	0.048	
3 internships	0.440	0.580	0.239	1.553	0.048	
4 or more internships	0.302	0.748	0.585	1.353	0.033	
Academic Achievement Indicators						
Cumulative Undergraduate GPA	2.474	4.066	0.000	11.868	0.289	***
Cumulative SAT Score (Math & Verbal)	-0.001	0.001	0.275	0.999	0.000	

Number of Observations = 1,062

Pseudo R-square = 0.1506

Significance: *** p < 0.001; ** p < 0.01; * p < 0.05; ~ p < 0.10

Comparison groups: Female, White, Zero Internships

Model also includes fixed effects by year

Acceptance to First Choice Graduate/Professional School

While most students would like to know what it takes to get into graduate school, what they would probably like to know more are the factors that contribute to their acceptance to their first choice institution. Again, grades were a significant factor in predicting whether a student is accepted to his or her first choice institution. Holding other variables constant, a one-point increase in cumulative grade point average improves the odds of being accepted to one's first choice institution by a factor of 3.130. The baseline academic achievement measure, cumulative SAT score, was significantly and negatively associated with acceptance to one's first choice graduate school. Practically, a 100-point increase in SAT score decreases the probability of acceptance to one's first choice graduate school by about 6 percent (delta-p=-0.059). A student's intended field of study also mattered. Students who applied to Arts and Humanities, Engineering, Natural Science, and Other Health programs were significantly more likely to be accepted to their first choice institution, keeping all other variables constant. Students intending to go to law school were significantly less likely to be accepted to their first choice law school. Among demographic characteristics, gender was not a significant predictor of acceptance to one's first choice institution, though race was a significant predictor. Controlling for a variety of factors, Black students who applied graduate or professional school were less than half as likely to get into their first choice institution, compared to White students. Asian students were marginally less likely than White students to be accepted to their first choice institutions. Similarly, those with majors in the Natural Sciences were marginally less likely to be accepted. The only co-curricular experience that was significantly associated with acceptance to one's first choice institution was study abroad. Students who engaged in study abroad were more likely to be accepted to their first choice graduate or professional school by a factor of 1.426.

Table 7. Logit Regression of Acceptance to First Choice Graduate/Professional School Among Students who Applied to Graduate/Professional School

	Coef.	Std. Error	p	Odds Ratio	Delta P	Sig.
Student Characteristics						
Male	-0.221	0.116	0.127	0.802	-0.054	
Asian	-0.359	0.139	0.071	0.698	-0.087	~
Black	-0.776	0.161	0.026	0.460	-0.187	*
Hispanic	-0.419	0.205	0.179	0.657	-0.102	
Foreign/International	-0.062	0.377	0.877	0.940	-0.015	
Other	0.341	0.726	0.510	1.406	0.083	
Major Category						
Arts and Humanities	-0.121	0.181	0.554	0.886	-0.029	
Engineering	-0.454	0.181	0.111	0.635	-0.110	
Interdisciplinary	-0.055	0.234	0.823	0.946	-0.013	
Natural Science	-0.434	0.160	0.079	0.648	-0.105	~
Social Science	-0.116	0.188	0.583	0.891	-0.028	
Student Activities						
Independent Research	-0.088	0.141	0.567	0.916	-0.021	
Research with faculty	0.120	0.184	0.460	1.128	0.029	
Study abroad	0.355	0.217	0.020	1.426	0.086	*
Volunteer work	0.209	0.174	0.139	1.232	0.051	
1 internship	-0.094	0.162	0.597	0.910	-0.023	
2 internships	-0.267	0.144	0.154	0.766	-0.065	
3 internships	-0.343	0.168	0.147	0.710	-0.083	
4 or more internships	0.195	0.443	0.593	1.215	0.047	
Intended Field of Graduate/Professional Study						
Arts & Humanities	0.984	0.856	0.002	2.675	0.235	**
Engineering (including Architecture)	1.353	1.215	0.000	3.867	0.319	***
Law School	-0.536	0.131	0.017	0.585	-0.130	*
Medical School	0.235	0.333	0.373	1.265	0.057	
Natural Science	0.764	0.663	0.013	2.146	0.184	*
Other Health	1.379	1.190	0.000	3.970	0.324	***
Social Science	0.343	0.373	0.194	1.410	0.083	
Academic Achievement Indicators						
Cumulative Undergraduate GPA	1.141	0.814	0.000	3.130	0.271	***
Cumulative SAT Score (Math & Verbal)	-0.002	0.001	0.002	0.998	-0.001	**

Number of Observations = 1,062
Pseudo R-square = 0.0964
Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ~ $p < 0.10$
Comparison groups: Female, White, Zero Internships, "Other" Intended Field of Study
Model also includes fixed effects by year

Technical Details about Regression Models

I analyzed the fit of all three models with several post-estimation tests. Regarding the quality of the models, the post-estimation tests yielded mixed results. The Hosmer-Lemeshow (H-L) goodness-of-fit test provided evidence in support of each model. The job attainment model yielded an H-L statistic of 3.69 and a p-value of 0.8839. The graduate school acceptance model produced an H-L statistic of 6.01 and a p-value of 0.6466. Finally, the H-L statistic for the first choice acceptance model was 10.00 with a p-value of 0.2650. All three of these H-L tests indicate that the models fit the data. Results produced from the model classification tables were less encouraging. When constructing classification tables, the cut-values used were based on the mean percentage each outcome was achieved (0.3546 for the job attainment outcome; 0.8757 for acceptance to any graduate/professional school; 0.5791 for acceptance to first choice graduate school). The job attainment and first choice models, in particular, had fairly low correct classification rates (62.6% and 67.23%, respectively), while the graduate school acceptance model had a correct classification rate that was fair (72.5%). Another post-estimation test I used was generating a Receiver Operating Characteristics (ROC) Curve for each model. The area under the curve, also known as the “c” statistic, can range from 0.5 (no predictive accuracy) to 1 (perfect predictive power). For the graduate school and first choice acceptance models, the “c” statistic were 0.7612 and 0.7075, respectively, which are fair values. The job attainment model, however, generated a value that would be considered poor (0.6786). These low correct classification rates and “c” statistics suggest that additional model refinement is needed for the job attainment model.

Limitations

While this study provides institutional guidance on achievement of post-baccalaureate outcomes, it also has some limitations that should be acknowledged. First, it is purely correlational. There will be no grounds for making any causal inferences based on the results. There is no opportunity for randomization nor does the data lend itself well to any of the econometric techniques that can be used. Second, the institution is a highly selective private institution with, for the most part, a bright, motivated student body. The findings could only be generalized to institutions with similar student profiles, curricula, geographies, and institutional missions. If one were to be truly cautious, he or she would not generalize beyond the Tufts campus. Third, the survey captures students at a single moment in time. While a student might not have a job at the time of the survey in late April or early May, he or she might have attained one only a week later. Additionally, students who started their job search 6 months prior to the survey are treated the same as those students who might have started their job search the day before they responded. Fourth, there is a presumption that an intended outcome of a Tufts education is to find a job or to get into graduate school. This is certainly not the case. Tufts aims to produce well-rounded leaders and global citizens, not good employees. Finally, the study is limited due to variables that are unavailable. For the graduate school admissions models, it is a limitation that graduate admissions test scores are unavailable, as they are likely to be strong predictors of success. While one could assume that SAT scores are correlated with graduate school admissions tests, it is a considerable weakness that this variable is not included. It would also be useful to have student financial indicators. Also, not knowing the selectivity or even the number of institutions to which students apply produces a gap in the study. Family income and student debt load might also be helpful indicators for understanding student backgrounds and priorities for post-baccalaureate plans. Fortifying the dataset with these additional variables would also likely require adding additional cases to avoid cell sizes too small to run these models.

Discussion

This research may prove to be helpful for the Career Services staff at Tufts University and similar institutions. The regression results help identify factors that are significantly related to students' success in finding jobs and gaining admission to graduate school. Additionally, predicted probabilities tables can act as a guide when trying to understand a student's likelihood of achieving his or her post-baccalaureate goals. Beyond the practical implications of the research, it is my hope that the scholarly body of work in this area will be expanded. The achievement of post-baccalaureate outcomes is a key issue for students, their families, and institutions. Considering its importance, the dearth of research on this topic is surprising. While there is substantial body of work on the economic returns to education, specifically regarding income, research aimed at understanding outcomes as simple as whether or not a student has a job at the time of graduation and the factors contributing to graduate school acceptance is sorely lacking.

Regarding the results themselves, the findings regarding the relationship between academic major and job attainment were in line with conventional beliefs about preparation for the workforce. Students in Engineering, who developed a specific skill set, were more likely to have accepted a job, while those in "softer" majors in the Arts and Humanities were less likely to have a job at the time of the survey. Similarly, the type of position one chooses to pursue proved to be a significant indicator of success, presumably due in part to different fields of employment working on varying recruiting cycles. As one would expect, the experience of internships provided students with exposure to the workforce and, as a result, students who participated in internships were significantly more likely to have accepted a position. Interestingly, this finding applied only to students who had more than one internship. Students who participated in only one internship were no more likely than their classmates who did not participate in any internships to have accepted a job by graduation. It was mildly discouraging to see that participation in other co-curricular activities had little bearing on a student's success finding a job.

Considering acceptance to graduate school, the results confirmed the importance of performing well academically if a student wants to improve his or her chances of getting into graduate or professional school. According to the acceptance to any graduate school model, grade point average was nearly the only thing that mattered (with participation in study abroad being the only other significant predictor). Grades were also a significant positive predictor in the first choice model, while this model also demonstrated the importance of factoring for academic field. It was surprising that neither research activity, neither independent research nor research with faculty, was significantly related to graduate school admission. One would expect that participation in research would be an indicator of success in graduate school admissions. Participation in research suggests that a student might have a predisposition to activities in which a graduate student would engage. Additionally, participation in research might help him or her develop skills for graduate study. These activities, however, were not significantly associated with graduate school admission when controlling for other variables.

The findings related to gender raised a few questions that are worth exploring in future research. Holding other variables constant, males were significantly more likely to have found a job at the time of graduation. Notably, this takes into account controls for type of position following graduation. What explains this difference? It is possible that males and females start their job searches at different times. The job attainment model includes only students who had started their job search. However, of the total population of students in the dataset, females were more likely to have started their job search after graduation compared to males (16.6% and 12.6%, respectively). This difference could be indicative of females generally starting their job search later, even those who had started their job search prior to graduation. Females may also be drawn to higher quality or more selective positions. Of the students who had started their search at the time of the survey, females were slightly more likely to have been offered a position and refused (7.3% compared to 5.4% for males). Even with these explanations, the difference between the success of males and the success of females in attaining jobs was considerable and this finding has generated a compelling new research question.

The finding that Black students were significantly less successful in gaining admission to their first choice graduate school, compared to White students, was surprising and troublesome. Controlling for academic ability and field of study, one would expect graduate admissions processes to be equitable or to possibly favor underrepresented minority groups as an affirmative action measure. This is an instance in which it would be helpful to have access to some of the unavailable data mentioned in the limitations section. We do not know whether there are considerable differences in the institutions to which Black students and White students apply. It is possible that Black students are more likely to reach for their first choice school, while White students have safer choices as their first choices. The data do not allow us to make this determination. Similarly, the model does not account for differences in graduate school admissions test scores. If there was a difference between Black students and White students regarding their test scores, this difference is not considered. Exploring this topic further is, unfortunately, beyond the scope of this particular research paper. Like the gender finding in the job attainment model, this race-based finding in the first-choice model presents a compelling new research question that should be explored.

Conclusion

Developing an understanding of the factors associated with achievement of post-baccalaureate outcomes is vital in a time when a college education is a considerable investment of time and money and job prospects are relatively poor due to the weak economy. Demonstrating that students have a bright future beyond graduation is critical for institutions. In an outcomes-based era of accountability in postsecondary education, being able to explain how outcomes such as job attainment and graduate/professional school admission are achieved is valuable in answering to both internal (students, parents, faculty/staff, alumni) and external (governing boards, federal and state governments, boards of trustees) constituents. Such analyses can help career services and academic administrators develop strategic plans to help achieve these outcomes. While this study is not without its limitations, it can act as a starting point for similar institutional studies. An analysis such as this should be able to provide some useful information that can inform decision-making and highlight some additional needs for future research.

Literature

Becker, G. S. (1975). *Human Capital,* 2nd ed. New York: Columbia University Press.

Bellas, M. (2001). Investment in Higher Education: Do Labor Market Opportunities Differ by Age of Recent College Graduates? *Research in Higher Education, 42(1),* 1-25.

Bowen, W. G., Bok, D. C., & Shulman, J. L. (2000). *The shape of the river: long-term consequences of considering race in college and university admissions*. Princeton University Press.

Grayson, J. (1997) *Who gets jobs? Initial labour market experiences of York graduates* (Research report). Toronto: York University, Institute for Social Research.

Hearn, J. (1987). Impacts of undergraduate experiences on aspirations and plans for graduate and professional education. *Research in Higher Education, 27(2),* 119-141.

Mincer, J. (1993). *Studies in Human Capital*. Brookfield, VT: Edward Elgar Publishing.

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

Paulsen, M. (1998). Recent Research on the Economics of Attending College: Returns on Investment and Responsiveness to Price. *Research in Higher Education, 39(4),* 471-489.

Perna, L. W. (2004). Understanding the Decision to Enroll in Graduate School: Sex and Racial/ethnic Group Differences. *Journal of Higher Education, 75(5),* 487–528.

Perna, L. W. (2005). The benefits of higher education: Sex, racial/ethnic, and socioeconomic group differences. *The Review of Higher Education, 29(1),* 23–52.

Pryor, J. H., Hurtado, S., Saenz, V. B., Santos, J. L., & Korn, W. S. (2007). The American freshman: Forty year trends. Los Angeles: Higher Education Research Institute, UCLA.

Sagen, H., Dallam, J., & Laverty, J. (2000) Effects of Career Preparation Experiences on the Initial Employment Success of College Graduates. *Research in Higher Education, 41(6),* 753-767.

Stoeker, J. & Pascarella, E. (1991) Women's colleges and women's career attainments revisited. *Journal of Higher Education, 62(4),* 394-406.

Strayhorn, T. (2008) Influences on Labor Market Outcomes of African American College Graduates: A National Study. *Journal of Higher Education, 79(1),* 28-57.

U.S. Department of Education. (2002). *Digest of education statistics* (No. 92-097). Washington, DC: National Center for Education Statistics.

Zhang, L. (2005). Advance to Graduate Education: The Effect of College Quality and Undergraduate Majors. *Review of Higher Education, 28(3),* 313-338.

TRACKING STUDENT RETENTION WITH UPDATING INFORMATION

Tom Gutman
Senior Institutional Research Analyst
Office of Institutional Research
New York University

Introduction

Sociologists, institutional researchers, practitioners of higher education, economists, and statisticians have explored the problem of student retention at colleges and universities for over forty years. Some researchers have focused primarily on developing theories as to why students attrit (Tinto, 1975; Bean 1980). Some researchers examine how precollege characteristics and demographics influence retention behavior (Bettinger, et al.,2011; Murtaugh et al., 1999). DesJardins et al. (1999, 2002) and Singell (2003) examine the relationship between financial aid and student persistence.

There is a diverse set of statistical methods used to describe the relationship between multiple student characteristics and the likelihood of attrition. Some studies utilize structural equations modeling to test empirically a combination of the theories proposed by Tinto and Bean (Cabrera et al., 1999). Others use logistic regression to predict student retention within a fixed time period, such as – but certainly not limited to – retention to the second fall (Bettinger et al., 2011). A smaller number of studies employ a set of tools called survival analysis or event history analysis (Murtaugh et al., 1999; DesJardins et al, 1999, 2002). Survival analysis is particularly useful for examining time-to-event or time-to-failure problems, in which one is concerned with the questions of who experiences a particular event, when do they experience the event, and why.

This paper follows the approach used by DesJardins et al., (2002), in which the authors fit a multivariate Cox proportional hazards model to enrollment, financial aid, precollege, and demographic data for students attending the University of Minnesota. The focus of that article is to model the effects of changing financial aid awarding strategies on student retention, which the authors accomplish by using their model to project how students' survival probabilities differ under various financial aid specifications. This paper applies a similar concept, namely building a Cox proportional hazards model for the purposes of predicting individual students' attrition risks on a term-by-term basis throughout their careers. The predictive model serves the primary purpose of identifying who is most likely to attrit and elucidates why a particular student is at risk.

The Office of Institutional Research at New York University has previously conducted a retention study using logistic regression to predict retention of new freshmen to the second fall of enrollment. This study represents the first attempt to apply a survival analysis approach to examine student retention at this large, private research institution. It is the first step in a broad attempt to understand retention behavior of NYU's students, one that hopefully will be followed by concrete action to improve the retention rate of the institution. Student retention is a dynamic problem, one that requires a flexible method that can incorporate the most recent set of information, model how the effects of various predictors change over students' careers, and can be easily used to target at-risk students.

Methodology

A Description of Survival Analysis

Survival analysis is a set of statistical tools for the analysis of time-to-event or time-to-failure data. It is an approach most commonly found in biostatistics, though there are applications in economics and sociology as well. The concept is best explained through an example from biostatistics. There is a beginning point of a study; for example, patients are diagnosed with some disease and enter a clinical trial for a new medication; some patients receive the treatment, and some receive a placebo. The researchers attempt to determine how the treatment affects the likelihood and duration of survival, and to assess the magnitude of the difference in survival between those patients receiving the treatment and those receiving the placebo.

In the context of student retention, students enter a university as freshmen with a set of precollege characteristics, such as demographics and high school performance metrics. Before enrollment, some students receive offers of financial aid from the institution to reduce the cost of attendance. The institutional researcher, sociologist, economist, or statistician attempts to determine how one's precollege and financial characteristics affect the university's likelihood of retaining the new students; however, the set of relevant information changes throughout the student's career. For instance, after one semester, the student has a GPA. Perhaps a student's financial aid package changes contingent on certain circumstances. The researcher must integrate the new information into the analysis in order to gain the best insight into the outcome – whether the student ultimately decides to continue studying at the institution and, hopefully, graduate. The survival analysis techniques utilized herein to explore student retention include Kaplan-Meier estimates of the survival function and the Cox proportional hazards model, both of which are methods capable of incorporating updating information.

Kaplan-Meier Estimates of the Survival Function.

The Kaplan-Meier method allows one to estimate survival curves based on one or more predictive factors. A nonparametric approach, it provides an estimate of the survival probability across the duration of a study. These survival probabilities can be visualized by plotting them against time. A point on the plot at time t is the conditional probability of survival for a particular individual at time t given that the student is still in the study. In this study, Kaplan-Meier plots are used purely for data visualization, although it is possible to use the log-rank test to determine statistically significant differences between survival curves. A sample annotated Kaplan-Meier plot is present in the appendix.

Cox proportional hazards model.

The Cox proportional hazards model allows the inclusion of time-dependent variables, a marked advantage over other methods. This model is a semi-parametric approach that does not depend on an underlying a specific survival distribution for estimation. The Cox model assumes proportional hazards, which states that the hazard rates for two individuals i and j are constant over time, and that the effect of a variable on the survival probability is the same regardless of the point in the study point. Essentially, the effect of a particular covariate is the same across individuals and the same across time. That said, the Cox model can be extended to allow for changes in the effects of a variable across time; unfortunately, one must specify the functional form that temporal effects take. For example, if one wants to allow a variable's effect to change linearly, one would specify the model as

$$y = \beta_1 X + \beta_2 (Xt) \tag{1}$$

for a given covariate X. Specifying functional time dependence allows one to circumvent the proportional hazards assumption in cases when the data do not display proportional hazards.

It is also possible to estimate stratified Cox models, in which one can estimate separate specifications of a Cox model for strata of a particular continuous or categorical predictor. A stratified Cox model allows for heterogeneity in the baseline hazard function; that is, the model allows for different survival probabilities for students in different strata (Kleinbaum & Klein, 2005). This is particularly useful when a predictor violates the proportional hazards assumption. Unfortunately, when one uses a stratified Cox model, one loses the ability to calculate explicitly the effect of the stratified variable. This is acceptable if the stratified variable is not the primary focus of the study, and the procedure does not affect the ability to predict hazard rates in new data.

Censoring.

Individuals in time-to-event studies either experience the event in question in the specified study time, or are said to be censored. Censoring implies that some of the data are incomplete. For instance, a student who graduates at the end of her eighth term has eight terms worth of data without attrition occurring; this student exits the study after the term of graduation and is considered censored. This is consistent with the definition of censoring in DesJardins et al. (2002), which treats graduation as an event competing with first stopout, and considers students who graduate before the first stopout as censored at the term of graduation. Similarly, a student currently working toward his degree in the sixth term has six terms of data without attrition occurring. One can say that this student has not attrit from the institution yet, but that can change in future terms.

Data Description

The data consist of all first-time, full-time, baccalaureate-degree-seeking freshmen entering in the fall semesters of 2005, 2006, 2007, and 2008. All subsequent spring and fall semesters in which a student is in attendance are captured through the fall semester of 2010. For the purposes of testing the model, the Fall 2010 term for the cohorts included in the study was separated from the rest of the data; thus, Fall 2010 data were not used in the estimation of the predictive model. There are no summer terms included in the data set, as the enrollment is not substantial enough to merit their inclusion. There are a total of 13,248 total freshmen with valid data, translating to a total of 65,628 rows; each row in the data represents a single student and term combination.

For each term for every student, the data include a student's

- geographic origin, classified into four categories (New York City, other New York State, other United States, and international);
- SAT score, or where appropriate, the converted ACT score;
- college GPA at census date;
- institutional aid (hereafter called "inside aid"), which includes any grant or scholarship funded by the university;
- loans, aggregated together regardless of the type, size, or interest rate;

- "outside aid," which includes any external private aid of which the university has knowledge and Pell, TAP, or other federal funds;
- work study funds;
- financial need, assumed to be zero if a student did not file the FAFSA.

For several reasons, the first term of data for all students was omitted. There were very few students with a valid first term college GPA. In effect, GPA becomes a lagged variable, which eliminates all students' first terms. Additionally, DesJardins et al. note the importance of using financial aid offers rather than the amount of financial aid accepted in order to avoid potential causation problems; otherwise, it becomes impossible to determine if a student leaves because his aid was revoked or if the student did not accept any aid because he is no longer attending the university. Unfortunately, utilizing financial aid offers is insufficient with this data set, as it was common data practice at NYU to change offer amounts to zero if a student did not accept the offer. To circumvent this problem, all of the financial variables were lagged by one term.

Consistent with DesJardins, et al. (2002), the dependent variable captures the timing of the first stopout. All data are captured as of the census date for each term, which occurs during the third week after the beginning of classes. Using this timing for the data does not seem initially intuitive, as the attrition variable does not capture students who exit the university after the census date; instead, the data record these students as attrit in the following terms. Because all data are lagged by one term, however, this definition for attrition is convenient. To clarify, suppose that a student attrits in October of her first term. At census date of term 2, she is recorded as an attrit student, and all of the predictors contained in that particular row of data correspond to her *first term* of attendance. Given that the student attrit in the first term, one can justify this timing for the attrition variable.

Table 1 presents the number of students in attnendance, the number of attritted students, and the number of censored students by the enrollment period. There are 2,824 students who are censored after four terms. This group consists primarily of the students who entered in fall 2008, for whom there are only four terms of data in the data set. Recall that censored students include graduates as well.

Table 1: Number of students present in the data by number of terms attended

Enrollment Period	Number of Students	Number Attrit	Number Censored	Cumulative Number Attrit	Survival Rate[1]
2nd Semester	13247	300	0	300	97.7
3rd Semester	12947	805	0	1105	91.7
4th Semester	12142	322	2824	1427	89.2
5th Semester	8995	363	9	1790	85.6
6th Semester	8623	178	3253	1968	83.9
7th Semester	5193	125	768	2093	81.8
8th Semester	4300	109	4050	2202	79.8
9th Semester	141	68	34	2270	41.3
10th Semester	39	15	24	2285	25.4

[1] This is the Kaplan-Meier estimate of the survival rate. A detailed example of how to calculate this is provided in the appendix.

Figure 1 displays the average lagged financial aid offer by type of aid. The average offer for each type of aid is consistent over the enrollment periods until the eighth term of enrollment (denoted in the plot at term 9 because of the one term lag), after which the average aid offer plummets. Figure 2 displays the number of aid offers by the type of aid. The number of aid recipients slowly falls over time as students attrit and are censored from the data. Initially, however, approximately 60% of the students in their second term are receiving some form of financial aid. By the ninth term of enrollment, very few students are offered aid. This occurs because financial aid packages at NYU expire after four years; while it is possible to appeal and extend the financial aid package, this does not happen often in practice.

Figure 1: Average financial aid offer by type of aid

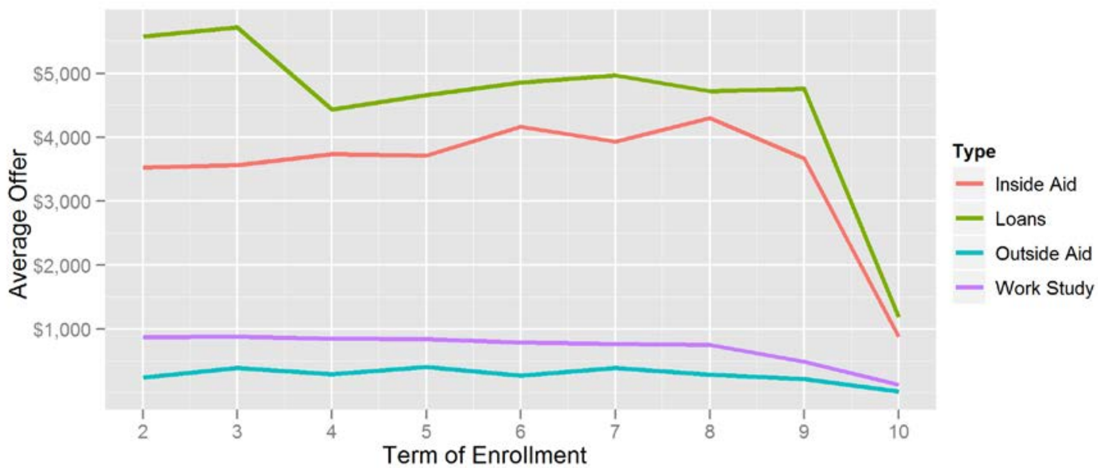
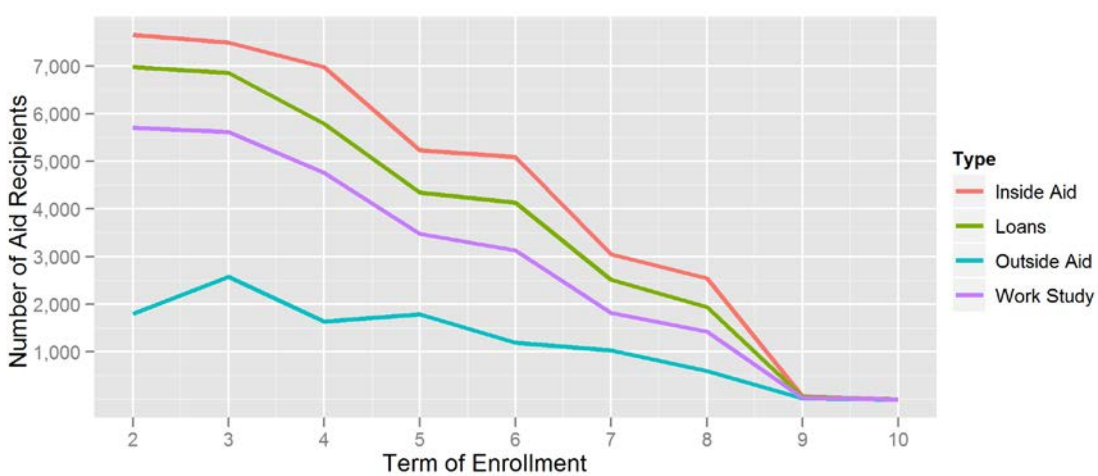


Figure 2: Number of Aid Recipients by type of aid

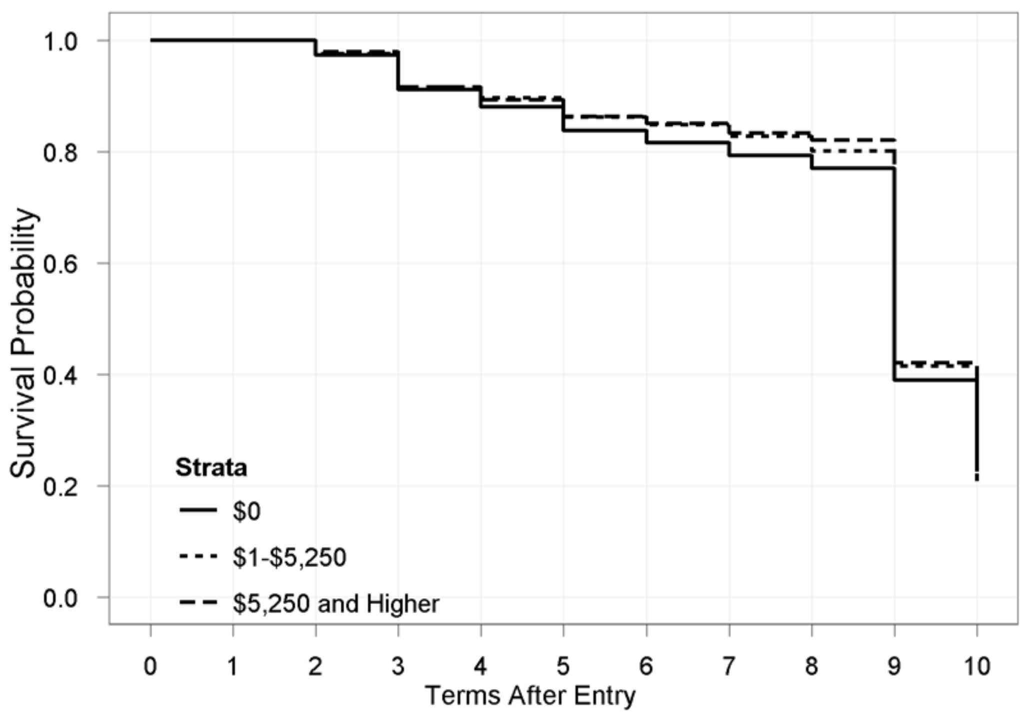


Exploring the Survival Distribution

For data exploration, Kaplan-Meier plots were constructed for every predictive variable utilized in the study. For brevity, there are only three plots provided – the survival curves segmented by inside aid, by financial need, and by college GPA. The remaining Kaplan-Meier plots for the other covariates are available upon request. These plots aided in the decision to include particular covariates; inclusion depended on sufficient spread between the survival curves for categories of a variable. It is important to note, however, that because these plots are univariate analyses one must use caution in making inference about the impact of a particular predictor on the survival probability.

Figure 3 displays the Kaplan-Meier plot of the survival distribution segmented by inside aid. Approximately 40% of students receive $0 worth of inside aid. The remaining two categories were established by taking the value that split the nonzero portion of the aid distribution in half. Because inside aid is a time-dependent variable (as are financial need and college GPA), students flow in and out of these categories across time. The $0 stratum has a survival curve consistently lower than the survival curves of the other two strata, containing the aid recipients. This plot suggests that there is a positive relationship between inside aid and the survival probability.

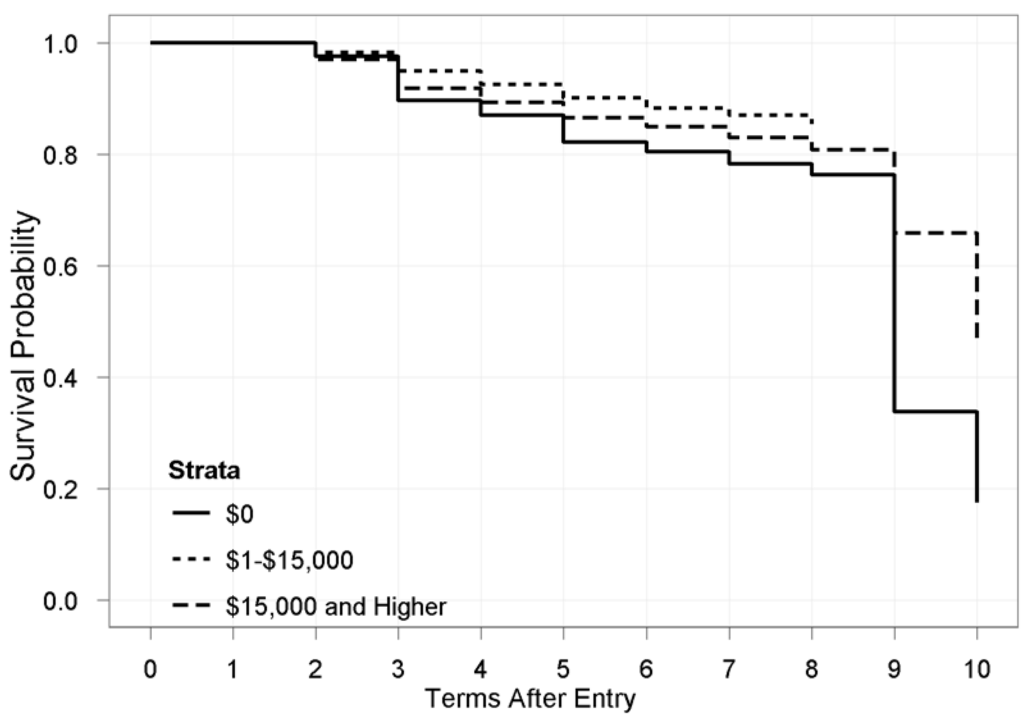
Figure 3: Kaplan-Meier Plot of Survival Curves by Inside Aid



The Kaplan-Meier plot of the survival distribution segmented by financial need is provided in Figure 4. Need was categorized in a way similar to the way in which inside aid was categorized. The separation of the survival curves suggests a positive relationship between financial need and the survival probability, though the survival curve for students in the highest need stratum is lower than the survival curve for students in the lower (nonzero) need stratum.

Figure 5 presents the Kaplan-Meier plot of the survival distribution segmented by college GPA. The four categories represent the quartiles of the college GPA distribution for the entire data set. There is a large negative disparity between the survival curve capturing students in the lowest GPA quartile and the three curves representing students in the upper three quartiles. It is important to caution against interpreting a GPA change from 3.23 to 3.22 – the separation between the bottom of the lower-middle quartile and the top of the lowest quartile – as a major change in the survival probability for these groups. Categorization strips away a considerable amount of information; nonetheless, this relationship between the survival probability and GPA is striking in magnitude, though not entirely surprising.

Figure 4: Kaplan-Meier Plot of Survival Curves by Financial Need



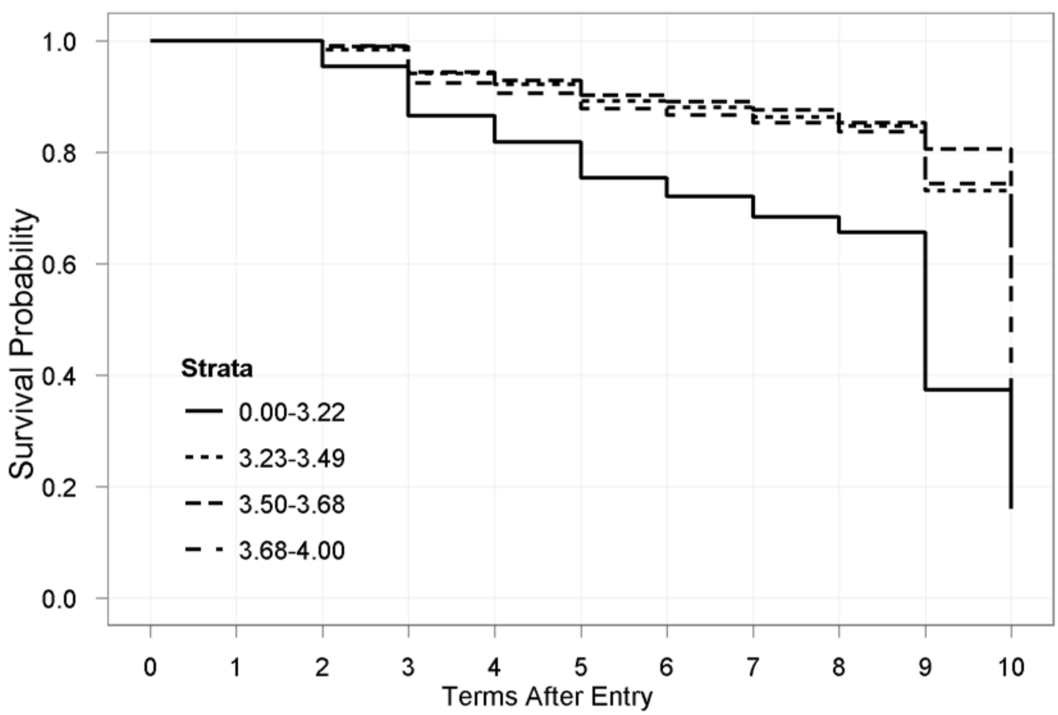
Results

Constructing the Cox Proportional Hazards Model

Model selection was guided by examining Kaplan-Meier plots, and models were fit in the forwards and backwards directions. The backwards model selection process – starting from an unrestricted model and testing simpler versions of the model – is detailed in Table 2. Models are judged on the basis of their log likelihood. A method to determine significant differences in log likelihood between two models is the likelihood ratio test (LR). The test statistic is:

$$2(Log\ Likelihood_{Unrestricted} - Log\ Likelihood_{Restricted})$$

Figure 5: Kaplan-Meier Plot of Survival Curves by College GPA



which is χ^2 distributed with degrees of freedom equal to the difference in the number of parameters of the unrestricted model and the number of parameters of the restricted model. If the p-value generated by this test is less than 0.005[2], then the log likelihood of the unrestricted model is statistically significantly higher than the log likelihood of restricted model, indicating better fit. Restricted models with statistically significantly lower log likelihood are denoted by an asterisk. It is important to note that this is merely one criterion for model selection, however.

The unrestricted model – the baseline – is stratified by cohort, by geographic origin, and by a binary variable indicating whether or not the student filed the FAFSA. The model includes SAT scores, college GPA at census, inside aid, outside aid, loans, work study, and financial need. Linear time dependence was specified for all covariates. All of the financial variables have been normalized to $1,000 changes. The unrestricted model has statistically significantly higher log likelihood at the 0.005 level than most of the alternative specifications tested. The exceptions include the specification excluding inside aid, the specification omitting outside aid, and the specification excluding linear time effects for the financial aid variables. The specification omitting all financial aid variables had significantly lower log likelihood than the unrestricted model, suggesting that the financial aid variables are jointly significant.

[2] Why 0.005? There are ten models tested; thus, one must adjust the standard 0.05 significance threshold by dividing by ten.

Table 2: Model Selection Table

	Log Likelihood	Chi-Squared Statistic	Degrees of Freedom	Probability Value
Unrestricted Model	-12960.7	--	--	--
Time Independent Model	-12980.5	39.6	6	< 0.0001*
SAT Scores Omitted	-12970.6	19.73	2	0.0001*
College GPA Omitted	-13924.9	1928.43	2	< 0.0001*
Inside Aid Omitted	-12963.8	6.17	2	0.0458
Loans Omitted	-12968.2	14.83	2	0.0006*
Outside Aid Omitted	-12962.3	3.12	2	0.2106
Work Study Omitted	-12967.1	12.7	2	0.0018*
Linear Time Effects for Financial Aid Variables Omitted	-12966.2	10.99	4	0.0267
No Financial Aid Variables	-12977.2	33.03	8	0.0001*
No Financial Need	-12966.9	12.29	2	0.0021*

DesJardins et al. (2002) argues that students respond to a set of prices and subsidies, and thus aid variables should be modeled together. Additionally, the model omitting linear time effects is slightly less flexible than the unrestricted model. For these reasons, the unrestricted model is favored over all alternative specifications.

Table 3 contains the logistic regression results for unrestricted model. The main effect and the time effect should be interpreted in tandem. The coefficient on the time effect depends on the number of terms that have passed (recall equation 1). For example, for the third term, the total effect of a variable is the sum of the main effect and the three times the time effect. Coefficients greater than zero imply higher attrition risk and vice versa. When the main effect has the same sign as the time effect for a given covariate, the total effect of that covariate strengthens over time. If the signs of the main effect and time effect for a given predictor are opposites, the total effect of the predictor weakens over time. In the case of SAT scores, the total effect initially suggests lower attrition risk; however, over the student's career, the sign of the total effect flips from negative to positive, implying higher attrition risk. Finally, the exponential function of a particular coefficient represents the hazard ratio. The interpretation is similar: values above one imply higher attrition risk, while values below one imply lower attrition risk.

The most notable effect from the model is the total effect of college GPA. The coefficient is normalized to a one-point increase in GPA – an extreme change. Even if one normalizes this effect to a 0.1 point change in GPA, the effect on the likelihood of attrition is strong; additionally, the impact of college GPA gets considerably stronger over the career of a student. The total effect of inside aid initially predicts higher attrition risk, though after the fifth term, the sign of the total effect changes. The same is true for loans.

Table 3: Parameter Estimates from the Unrestricted Model

	Coefficient	exp(Coefficient)	Robust SE	Z Score	Pr(>|z|)[3]
SAT Score	-0.000500	0.999	0.000560	-0.905	0.36538
SAT Score - Linear Time	0.000312	1.000	0.000127	2.421	0.01549
GPA at Census	-0.831140	0.436	0.077898	-9.848	< 0.00001*
GPA at Census - Linear Time	-0.100300	0.905	0.021535	-4.343	0.00001*
Need/$1,000	0.021355	1.022	0.011419	1.790	0.07344
Need/$1,000 - Linear Time	-0.001320	0.999	0.002585	-0.499	0.61774
Inside Aid/$1,000	0.042899	1.044	0.017239	2.413	0.01583
Inside Aid/$1,000 - Linear Time	-0.008320	0.992	0.003659	-2.117	0.03424
Loans/$1,000	0.029711	1.030	0.008830	3.175	0.00150*
Loans/$1,000 - Linear Time	-0.004580	0.995	0.001964	-2.210	0.02708
Outside Aid/$1,000	0.044840	1.046	0.081107	0.536	0.59220
Outside Aid/$1,000 - Linear Time	0.002023	1.002	0.018447	0.103	0.91768
Work Study/$1,000	-0.061980	0.940	0.072525	-0.782	0.43432
Work Study/$1,000 - Linear Time	-0.007700	0.992	0.015532	-0.465	0.64198

Using the Model to Predict Outcomes in New Data

For the purposes of prediction, the results from the unrestricted model were applied to a different set of data, namely the Fall 2010 term for the same cohorts. To gain a metric for risk, there are two figures that can be calculated from the model for each individual student: the linear predictor and the risk score. For each student, k, the linear predictor is

$$Linear\ Predictor = \sum_{j=1}^{J} \beta_j X_{jk} \quad (2)$$

for all J predictors. The risk score is simply the exponential function of the linear predictor, or

$$Risk = e^{Linear\ Predictor} \quad (3)$$

Figure 6 provides a histogram of the bottom 99% of the predicted risk score distribution after calculating the risk scores for every student in the new data. Scores in the top 1% of the risk distribution were quite extreme and therefore were not shown (the maximum risk score calculated was 144; 99% of the distribution lies between 0 and 5).

Whether or not students attrit in Fall 2010 is already known. Table 4 presents a comparison between the predicted risk and the percentage of students who attritted in each category of the risk distribution. In this table, risk scores are separated into deciles. In addition to the risk scores, the average GPA, average need, and average inside aid by risk decile are provided. Students in the top decile of predicted risk have a dramatically

[3] The asterisk indicates statistical significance at the 0.05/14 = .0035, where 14 is the number of parameters in the model.

higher attrition rate than students in the other nine deciles; three times as many students attritted in this category compared to the next lowest decile.

Figure 6: Histogram of Bottom 99% of Predicted Risk Distribution for Fall 2010 Data

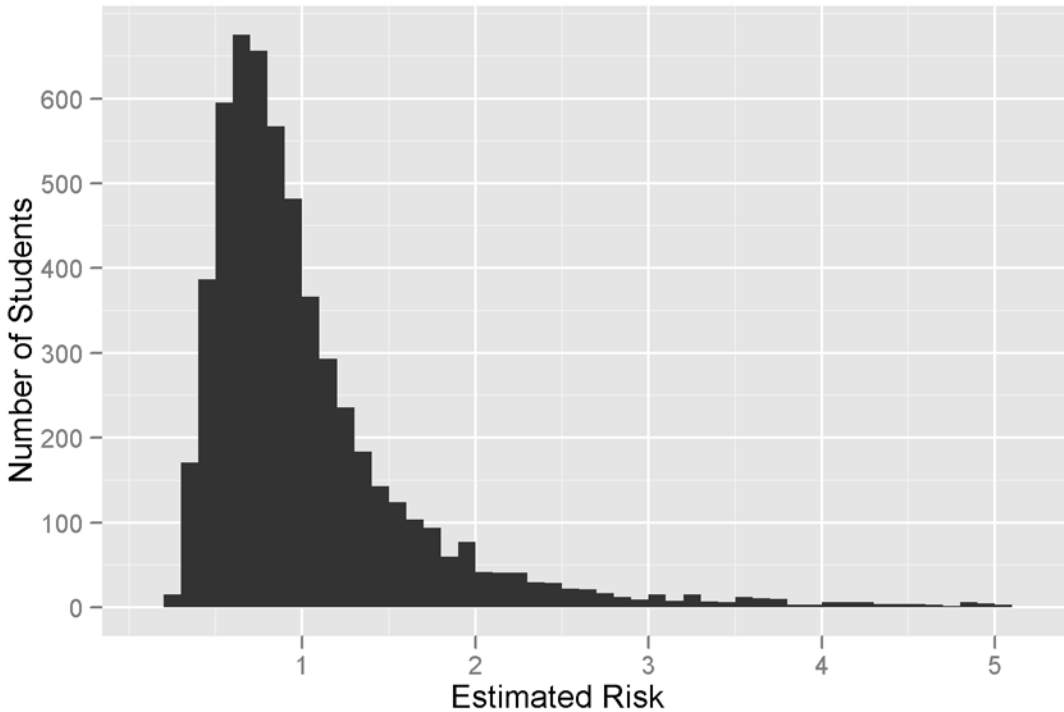


Table 4: Actual Attrition Rates by Deciles of Predicted Risk

Decile of Predicted Risk	N Students	Attrition Rate	Average GPA	Average Need	Average Inside Aid
0%-9%	569	3.3%	3.77	$19,697	$8,597
10%-19%	569	3.5%	3.70	$17,529	$6,884
20%-29%	568	3.0%	3.68	$13,951	$5,962
30%-39%	569	2.5%	3.63	$12,565	$4,761
40%-49%	568	5.1%	3.57	$10,850	$4,262
50%-59%	569	2.3%	3.52	$9,057	$3,523
60%-69%	569	4.4%	3.42	$9,886	$3,929
70%-79%	568	5.6%	3.32	$8,750	$3,110
80%-89%	569	6.5%	3.18	$8,083	$3,342
90%-100%	569	17.8%	2.79	$6,221	$2,635

This result is quite encouraging and worthy of further exploration. Table 5 examines the top decile of the predicted risk distribution. The attrition rates on students in the top four percentiles of predicted risk are high. This provides a set of 228 students who can be targeted for intervention; additionally, it is possible to examine these students at an individual level. This raises the possibility of creating an intervention plan tailored to individual students, an idea possessing much potential.

Table 5: Actual Attrition Rates of the Top 10% of Predicted Risk

Percentile of Predicted Risk	N Students	Attrition Rate	Average GPA	Average Need	Average Inside Aid
90%-90.9%	57	17.5%	3.04	$6,803	$2,604
91%-91.9%	57	7.0%	3.05	$6,860	$2,788
92%-92.9%	56	14.3%	2.98	$9,074	$3,061
93%-93.9%	57	12.3%	2.96	$5,804	$2,250
94%-94.9%	57	8.8%	2.87	$6,860	$3,516
95%-95.9%	57	12.3%	2.88	$6,099	$1,756
96%-96.9%	57	17.5%	2.74	$6,041	$3,257
97%-97.9%	57	19.3%	2.69	$5,043	$2,395
98%-98.9%	57	26.3%	2.51	$6,508	$3,654
99%-100%	57	42.1%	2.15	$3,168	$1,076

Conclusions

Survival analysis offers the best possible way to model student retention behavior. The techniques utilized herein allow one to take advantage of the fact that data constantly change throughout students' careers. A Cox model can be used to estimate the attrition probability in the middle of a student's time at the college or university. There are a number of ways that one can improve upon the framework of this model, such as the utilization of bootstrapping or cross-validation to validate the model results, and the inclusion of other predictors, such as ethnicity and high school GPA. The statistics, of course, are merely a small portion of the solution.

Student retention cannot be increased through the use of statistics alone. Effective interventions must be made if a student is struggling academically. Financial aid packages must be designed in such a way that the student is not over encumbered by loan debt and therefore feels pressure to exit postsecondary education early. Administrators require the ability to intervene and assist a student whose family's finances suddenly collapse. Counseling must be known to be available for students far from home who are struggling to adapt to a new environment. The statistical models do not accomplish any of those things alone. They do, however, shed light on the problem, a first step from which administrators can design an intervention plan. Indeed, statistical modeling of a problem and the resulting action plan are symbiotic; as an intervention plan progresses, administrators can call upon statistics to evaluate the impact of the intervention. Thus, the model used needs to be robust and flexible. Survival analysis provides that opportunity.

Appendix

Calculating Kaplan-Meier Estimates of the Survival Rate

The Kaplan-Meier method of calculating the survival rate for a particular time period accounts of censoring when applicable. For example, from Table 1, the calculation of the survival rate for term two (the first row) is simply:

$$2^{nd} \text{ Term Survival Rate} = (13247 – 300)/13247 = .977$$

For the third term, the survival rate is:

$$3^{rd} \text{ Term Survival Rate} = (13247 – 300)/13247 * (12947 – 805)/12947 = .917$$

The fourth term's survival rate is:

$$4^{th} \text{ Term Survival Rate} =$$
$$(13247 – 300)/13247 * (12947 – 805)/12947 * (12142 – 322)/12142 = .892$$
Or
$$3^{rd} \text{ Term Survival Rate}*(12142 – 322)/12142$$

The calculation of the fifth term survival rate has to take into account the fact that 2,824 students were censored at the end of the fourth term. This calculation is:

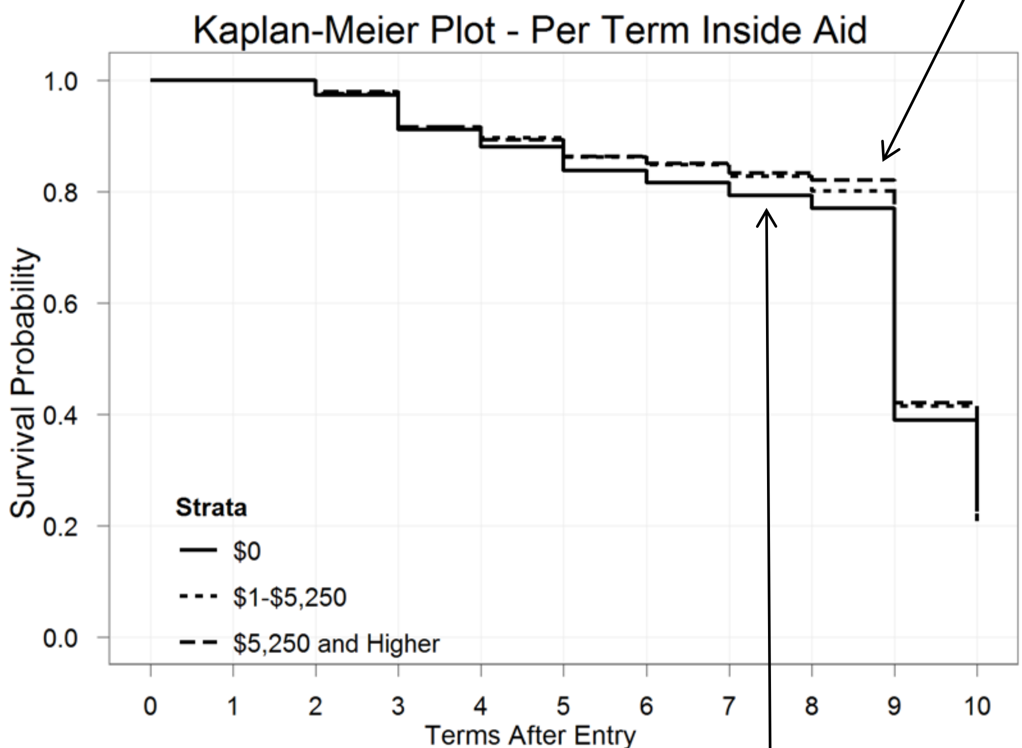
$$5^{th} \text{ Term Survival Rate} =$$
$$4^{th} \text{ Term Survival Rate}*(8995 – 365)/8995 = .856$$

Using 8,995 in the calculation reflects the fact that students were censored between terms 4 and 5. A general formula for the survival rate in a given term t is:

$$\text{Survival Rate}_t = \text{Survival Rate}_{t-1} * (\text{N Students}_t – \text{N Attrit}_t)/\text{N Students}_t$$

Annotated Kaplan-Meier Plot

The lines of the plot are the Kaplan-Meier estimates of the survival rate at a given point in time. Recall that the survival rate is a conditional probability.



Of primary interest is determining whether or not there is a difference between the survival curves of the different strata. Here, the curve for the $0 stratum is lower than the curves for the other two strata.

References

Bean, John (1980). "Dropouts and Turnover: The Synthesis and Test of a Causal Model of Student Attrition." *Research in Higher Education*, 17, pp. 155-187.

Bettinger, Eric, Evans, Brent, and Pope, Devin (2011). "Improving College Performance the Easy Way: Unpacking the ACT Exam." NBER Working Paper 17119. National Bureau of Economic Research. Cambridge, MA.

Cabrera, Alberto, Nora, Amaury, and Castañeda, Maria (1993). "College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention." *The Journal of Higher Education*. Vol.64, No.2., pp.123-139.

Cox, David (1972). "Regression Models with Life Tables (with Discussion)." *Journal of the Royal Statistical Society*. Series B, 34, 187-220.

DesJardins, Stephen, Ahlburg, Dennis, and McCall, Brian (1999). "An Event History Model of Student Departure." *Economics of Education Review*. Vol. 18. pp. 375-390.

DesJardins, Stephen, Ahlburg, Dennis, and McCall, Brian (2002). "Simulating the Longitudinal Effects of Changes in Financial Aid on Student Departure from College." *The Journal of Human Resources.* Vol 37, No.3, pp 653-679.

Guo, Shenyang (2010). *Survival Analysis.* Oxford University Press. New York.

Harrell, Frank, Lee, Kerry, and Mark, Daniel (1996). "Multivariate Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors." *Statistics in Medicine.* Vol. 15, pp 361-387.

Hosmer, David, and Lemeshow, Stanley (1999). *Applied Survival Analysis*. John Wiley & Sons. New York.

Kleinbaum, David, and Klein, Mitchell (2005). "Survival Analysis: A Self-Learning Text." New York. Springer Science.

McCall, Brian (1994). "Testing the Proportional Hazards Assumption in the Presence of Unmeasured Heterogeneity." *Journal of Applied Econometrics*. Vol. 9, pp. 321-334.

Murtaugh, Paul, Burns, Leslie, and Schuster Jill (1999). "Predicting the Retention of University Students." *Research in Higher Education*. Vol 40, No. 3, pp 355-371.

Radcliffe, Peter, Huesman, Ronald, and Kellogg, John (2006). "Identifying Students at Risk: Utilizing Survival Analysis to Study Student Athlete Attrition." Paper presented at The National Symposium on Student Retention, Albuquerque, NM.

Reason, Robert (2009). "Student Variables that Predict Retention: Recent Research and New Developments." *NASPA Journal*, Vol. 46, No.3, pp 482-501.

Singell, Larry (2004). "Come and Stay a While: Does Financial Aid Effect [sic] Retention Conditioned on Enrollment at a Large Public University." *Economics of Education Review*. Vol. 23, pp. 459-471.

Tinto, Vincent (1975). "Dropout from Higher Education: a Theoretical Synthesis of Recent Research." *Review of Educational Research*, 45, pp. 89-125.

Yamaguchi, Kazuo (1991). *Event History Analysis.* Sage Publications. Newbury Park, California.