## Conference Planning Team

**President:**
HEATHER KELLY

**Program Chair:**
BRUCE SZELEST

**Associate Program Chair:**
CATHERINE ALVORD

**Pre-Conf Workshop Coordinator:**
NICOLE MARANO

**Best Paper Chair:**
MELANIE SULLIVAN

**Evaluation Coordinator:**
TERRY HIRSCH

**Newcomer/Mentor Coordinator:**
PAULA MAAS

**Poster Session Coordinator:**
PAULA MAAS

**Proposal Peer Review Coordinator:**
MARK ECKSTEIN

**Publications Coordinator:**
BETH FREDERICK

**Local Arrangements Chair:**
JACQUELINE ANDREWS

**Exhibitor Coordinator:**
GURVINDER KHANEJA

**AV Coordinator:**
NORA GALAMBOS

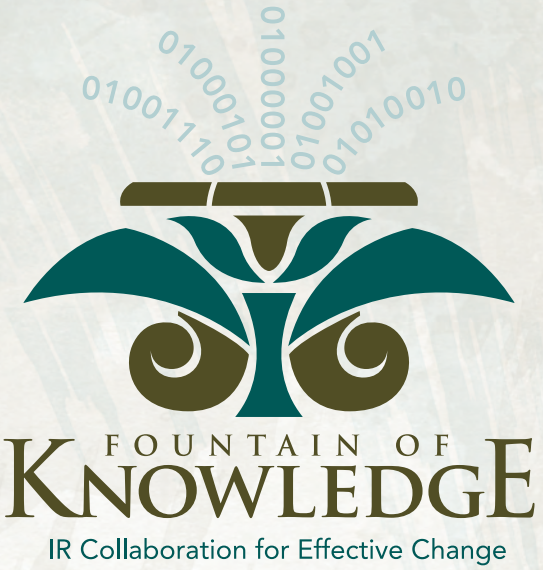
**Dine Around Coordinators:**
HIROSUKE HONDA &
KRISTINE ALTUCHER

**Local Arrangements Coordinator:**
PATRICIA FRANCIS

**Website Chair:**
MARK PALLADINO

**Conference Website Coordinator:**
CHRISTOPHER CHONCEK

**Administrative Coordinator:**
BETH SIMPSON




# 37TH ANNUAL CONFERENCE

## NOVEMBER 13 – 16, 2010

**The Saratoga Hilton, Saratoga Springs, New York**

## NEAIR OFFICERS & STEERING COMMITTEE

**Heather Kelly**
President

**Gayle Fink**
President – Elect

**Mitchell Nesler**
Past President

**Nancy Ludwig**
Recording Secretary

**Eleanor Swanson**
Treasurer

**Allison Walters**
Member-At-Large

**Mark Palladino**
Member-At-Large

**Ellen Boylan**
Member-At-Large

**Michael Dooris**
Member-At-Large

**Emily Dibble**
Member-At-Large

**Paula Maas**
Member-At-Large

**Bruce Szelest**
Program Chair

**Jacqueline Andrews**
Local Arrangements Chair

Dear Friends,

The 37th Annual NEAIR Conference held in Saratoga Springs, New York November 13-16, 2010 encouraged attendees to contribute to the *Fountain of Knowledge: IR Collaboration for Effective Change*. Three hundred conference attendees had the opportunity to share and gain invaluable information from institutional research and higher education colleagues. The 2010 Conference Proceedings is a result of the conference theme in action.

The Conference Program team led by Program Chair **Bruce Szelest** and Associate Program Chair **Cathy Alvord** developed a program filled with plenty of variety that included three plenary/keynote speakers, 15 contributed papers, 19 workshares, 14 techshares, 10 special interest groups, and four table topics. Poster Session Coordinator **Paula Maas** organized 14 posters to be on display. These offerings went through a blind peer review process facilitated by 57 proposal reviewers coordinated by **Mark Eckstein**. Pre-Conference Workshop Coordinator **Nicole Marano** organized 18 workshops with 199 participants. Exhibitor Coordinator **Gurvinder Khaneja** partnered with a record 20 exhibitors who offered 10 exhibitor showcases.

A big thanks goes to Publications Coordinator **Beth Frederick** for all her hard work and keen eye editing the Conference Program, as well as compiling and organizing the 2010 Conference Proceedings. The 2010 Conference Proceedings contains papers submitted by authors, as well as the 2010 Best Paper Award recipients. The award recipients were determined by Best Paper Chair **Melanie Sullivan** and her committee. The 2010 Best First Paper is **Joel Bloom's**, *"Issues in Web Surveys Student Populations: Response Rates and Post-Stratification Weighting."* The 2010 Best Paper is **Meredith Billings** and **Dawn Geronimo Terkla's**, *"Using SEM to Describe the Infusion of Civic Engagement in the Campus Culture."* The 2010 Best IR & Practitioner Report is John Runfeldt's, *"Organizing Student Tracker Results Using SPSS."* Poster Session Coordinator **Paula Maas** and her committee evaluated the poster displays to select **Marie Wilde**, for her poster titled *"Assessing Institutional Effectiveness Using a KPI Dashboard"*, as the 2010 Best Visual Display Award recipient.

Local Arrangements Chair **Jackie Andrews** and Local Arrangements Coordinator **Patty Francis** worked hard coordinating hotel, travel logistics and made sure we all enjoyed the local flavors (cupcakes, anyone?) and activities Saratoga Springs had to offer. AV Coordinator **Nora Galambos** assisted with technology and Dine Around Coordinators **Hirosuke Honda** and **Kris Altucher** made sure we were well-fed and had an additional networking opportunity.

Website Chair **Mark Palladino**, 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 **Terry Hirsch**.

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 2010 Conference Proceedings have to offer!

Wishing you all the best,
*Heather Kelly*
NEAIR President 2009-10

# Leadership & Committees 2009 – 2010

## Officers:

| | |
|---|---|
| President | Heather A. Kelly |
| President-Elect | Gayle Fink |
| Recording Secretary (’08-’11) | Nancy Ludwig |
| Treasurer (’08-’11) | Eleanor Swanson |

## Steering Committee Members:

| | |
|---|---|
| Past President | Mitch Nesler |
| Program Chair | Bruce Szelest |
| Local Arrangements Chair | Jackie Andrews |
| Member-At-Large (’07-’10) | Mark Palladino |
| Member-At-Large (’07-’10) | Allison Walters |
| 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 |
| **Administrative Coordinator (ex-officio)** | Beth Simpson |

## Standing Committees

### Program Committee

| | |
|---|---|
| **Chair** | **Bruce Szelest** |
| Associate Program Chair | Cathy Alvord |
| Pre-Conf Wrkshp Coord | Nicole Marano |
| Best Paper Coordinator | Melanie Sullivan |
| Evaluation Coordinator | Terry Hirsch |
| Newcomer/Mentor Coord. | Paula Maas |
| Peer Review Coordinator | Mark Eckstein |
| Poster Session Coord. | Paula Maas |
| Publications Coordinator | Beth Frederick |

### Professional Development Services Committee

| | |
|---|---|
| **Chair** | **Gayle Fink** |
| Member - Program Chair-Elect | Nicole Marano |
| Member (’08-’11) | Ellen Boylan |
| Member | Michelle Appel |
| Member | Paula Maas |
| Member | Linda Mallory |

### Technology Committee

| | |
|---|---|
| **Chair (’08 – ’10)** | **Mark Palladino** |
| Conf Web Coordinator | Annemarie Bartlett |
| Member | Chad May |
| Member | Chris Choncek |
| Member (ex officio) | Beth Simpson |
| Program Chair(ex officio) | Bruce Szelest |
| LAC (ex officio) | Jackie Andrews |
| **Listserv Manager** | **Fred Cohen** |

### Local Arrangements Committee

| | |
|---|---|
| **Chair** | **Jackie Andrews** |
| Exhibitor Coordinator | Gurvinder Khaneja |
| AV Coordinator | Nora Galambos |
| LAC Coordinator | Patty Francis |
| Dine Around Coords | Hirosuke Honda |
| | Kris Altucher |

### Membership Committee

| | |
|---|---|
| **Chair (’08 – ’10)** | **Allison Walters** |
| Chair-elect | Paula Maas |
| Member | Eileen McDonnell |
| Member | Laura Uerling |

### Site Selection Committee

| | |
|---|---|
| **Chair** | **Gayle Fink** |
| NEAIR Treasurer | Eleanor Swanson |
| Member | Michael Dooris |
| Member | Steve Thorpe |
| Member | Mark Palladino |
| Member | Jessica Shedd |

### Nomination Committee

| | |
|---|---|
| **Chair** | **Mitch Nesler** |
| 4 YR Public Sector | Braden Hosch |
| Member | Erin Ebersole |
| Member | Maren Hess |
| Member | Jeff Himmelberger |
| Member | Ingrid Skadberg |
| Member | Shannon Lichtinger |

NEAIR
North East Association for Institutional Research

# Leadership & Committees
# 2009 – 2010

## Ad Hoc Committees

**Finance Committee**

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

**Grants Committee**

| | |
|---|---|
| **Chair** | Michael Dooris |
| Chair-Elect | Emily Dibble |
| Member (2008 - 2011) | Cathy Alvord |
| Member (2007 – 2010) | Diane Petruccio |
| Member (2007 – 2010) | Ingrid Skadberg |
| Past Chair (ex officio) | George Rezendes |

## Conference Proposal Peer Reviewers

Mark Eckstein, Peer Review Coordinator

Govind Acharya
Lou Attinasi
Becky Bell
Felice Billups
Kim Bridgeo
Cristi Carson
Peggye Cohen
Ted Coladarci
Lisa Daniels
Cherry Danielson
Kathryn Doherty
Karen Durkin
Ricka Fine
Patty Francis
Michael Goldberg
Cheryl Goldstein
Lynne Govatos
Joan Graham
Marianne Guidos
Phil Handwerk
Matthew Hendrickson
Jill Hirt
Braden Hosch
Patricia Hulsebosch
Melanie Jenkins
Ethan Kolek
Arthur Kramer
Elisabeth Lackner
Melanie Larson
Mary Lynch
Chad May
Daniel McCollum
Peggy McKernan
Pat Mizak
Kevin Murphy
Louise Murray
Tiffany Parker
George Rezendes
Cheryl Rollins
Millie Savidge
Karen Schreiner
Jessica Shedd
Marian Sherwood
Yan Shi
Ingrid Skadberg
Jacki Stirn
Steve Thorpe
Shannon Tinney
Laura Uerling
Christopher Vinger
Lynn Wallace
Michael Whitcomb
Bill Wilson
Lillian Zhu

# Table of Contents



* ***Best First Paper Award - 2010***

◈ ***Best Paper Award - 2010***

## ACKNOWLEDGMENT

Contained within these pages of the NEAIR 37th Annual Conference Proceedings you will find eight contributed conference papers/presentations authored by eleven NEAIR colleagues.

Additional conference presentations are just a few clicks away—accessible within the NEAIR website "Members Only" section.

Special thanks go out to Heather Kelly, Beth Simpson, Melanie Sullivan, and Bruce Szelest for their guidance, patience, and support with all aspects of publication responsibilities during the course of this past year.

Beth E. Frederick
Publications Chair, 2010 – 2011
University of the Arts

# COLLABORATION BETWEEN STUDENT AFFAIRS & INSTITUTIONAL RESEARCH:

# A MODEL FOR THE SUCCESSFUL ASSESSMENT OF STUDENTS' COLLEGE EXPERIENCE

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

Joel D. Bloom, Ph.D.
Director of Assessment and Survey Research

University at Albany, SUNY

This paper was prepared for presentation at the 2010 Annual Meeting of the North East Association for Institutional Research, November 13-16 2010, Saratoga Springs, NY.

***ABSTRACT:*** *This paper highlights the collaborative efforts between Student Affairs and Institutional Research at the University at Albany by examining Student Affairs-Institutional Research collaboration at the University since 2007. Through collaboration, Student Affairs and Institutional Research have been able to convert assessment findings into effective change to enhance students' collegiate experience. These efforts include continuing and open dialogue, consultative scheduling and promotion of assessments, and the sharing of findings and outreach efforts to the campus community.*

## Introduction

In the summer of 2008, student affairs charged select personnel in each of the Division's 13 units with responsibility for unit-level assessment as well as introduced NASPA's Assessment Education Framework to "to assist those practitioners who have been charged with assessment to carefully and intentionally choose training options to support their assessment work" (NASPA Assessment Education Framework, 2009). In doing so, assessment in student affairs became a practical priority for professionals throughout the Division and expanded the professional network of assessment professionals for institutional research at the University. Furthermore, by assigning the task of assessment to a designated assessment professional and providing that professional with continuing education opportunities, student affairs was also able to utilize the data provided by institutional research to develop supporting assessment activities to inform decisions impacting improvements to student affairs programs, services and activities.

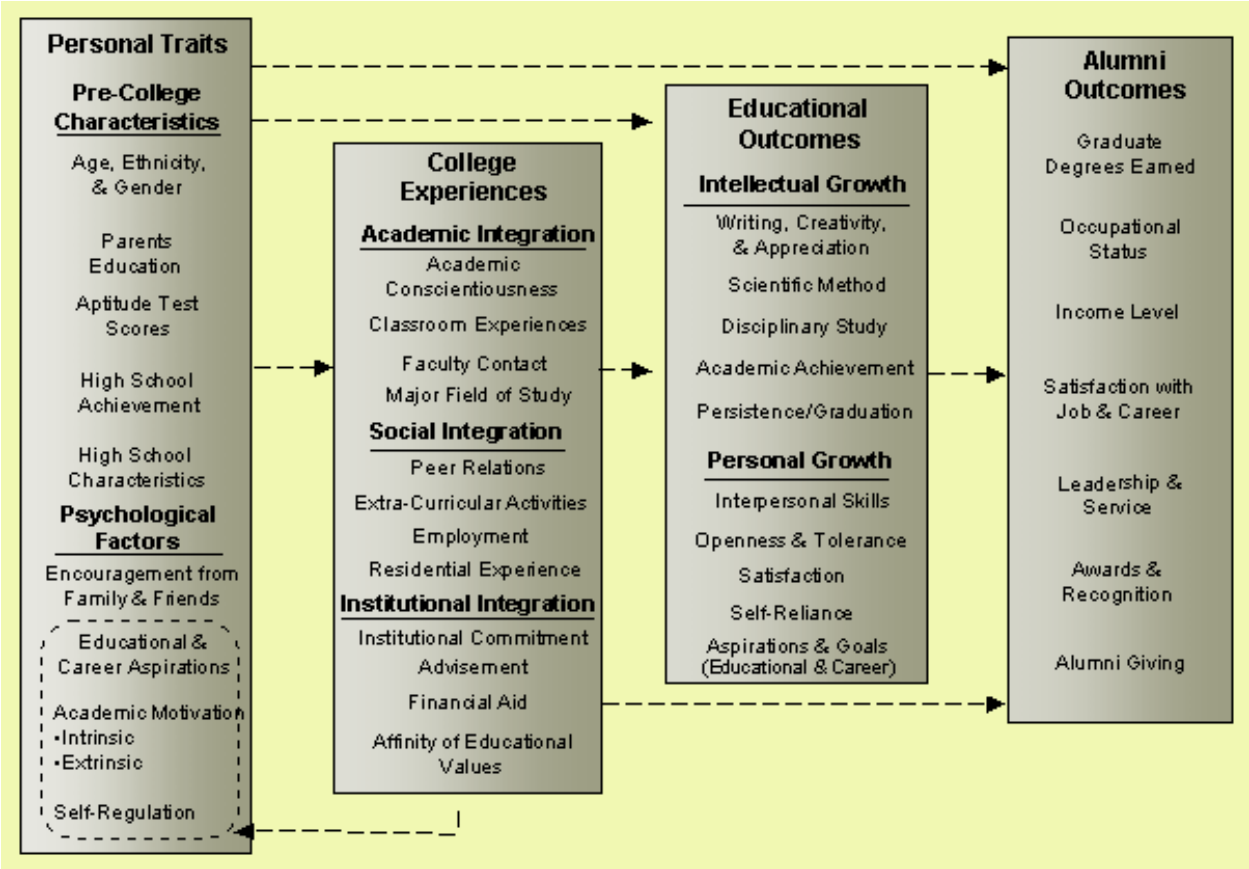
This move toward greater institutionalization of assessment activities within the Division did not occur in a vacuum; rather, it fit nicely into UAlbany's longstanding

philosophy of viewing the undergraduate experience – and assessment of it – as a coherent, integrated system. This view, summarized below in Figure 1, has come to be known as the Albany Outcomes Assessment Model. As described on UAlbany's assessment web page,

> The model… relates students' college experience to their pre-college characteristics, as depicted in the following chart. Findings from this research underscore the importance of connecting the classroom and related student experiences (e.g., academic, social) to student satisfaction and success. These assessment efforts, which have been conducted on a continuous basis by the Office of Institutional Research, have given the University a rich array of evaluative databases, including student opinion surveys, cohort studies, and alumni studies. (UAlbany 2)

The aspects of the model covered by the Division of Student Success fall largely under the second bar from the left, under "College Experiences/Social Integration," a category that includes peer relations, extra-curricular activities, employment and residential experiences.

Figure 1. The Albany Outcomes Assessment Model




Source: "The Albany Outcomes Assessment Model"
http://www.albany.edu/assessment/ualb_outcomes_model.html (UAlbany 1)

## Literature Review

Martin and Murphy (2000) suggest that collaboration can enhance the quality of students' educational experience and that successful partnerships put students at the center. Banta and Kuh (2000) identified "bureaucratic-structural barriers" due to an institution's "formal organizational arrangements" as an obstacle to collaboration in assessment. Like many units at today's colleges and universities, student affairs divisions often function in "silos" that limit meaningful collaboration with other units across campus, including institutional research offices. While institutional research offices

conduct a variety of assessments across the entire institution, student affairs' assessment of students' co-curricular experiences and satisfaction with various services largely occurs within the scope of the division or unit conducting the assessment and may not always be shared with institutional research offices and vice a versa.

As a result, "higher education leaders began to reexamine the need for integration of these roles and have advocated a change…from separatist to seamless" (Kezar, 2003, 137). During the last decade, collaboration in assessment has resulted in more seamless environments in which students have increased opportunities for learning in as well as out of the classroom as "connected experiences building upon each other" (Knefelkamp, 1991; Kuh, Douglas, Lund, & Gyurmek, 1994; Schroeder, C. S., & Hurst, J. C., 1996). Increased collaboration will better fulfill the institution's mission, improve retention and improve the total college experience for students (ACPA, 1994; Hyman, 1995; Kuh, 2006). ACPA's (1994) Student Learning Imperative indicated that "the more students are involved in a variety of activities inside and outside the classroom, the more they gain."

ACPA's (1994) Student Learning Imperative urges Student Affairs professionals to gather information to redesign policies and practices as well as evaluate programs and services to determine the degree to which they contribute to a student's undergraduate experience. Specifically, the Student Learning Imperative document concludes that "student affairs staff should participate in institution-wide efforts to assess student learning" (ACPA, 1994). Moxley (1999) suggests that, "student affairs divisions have employed a wide range of informal and formal structures for collecting information" but that the "research skills and interest of staff members, financial resources, existence of a campus research office, and the extent to which the chief student affairs

administrator…see student- and program-related information as a priority, all have an impact on the data collection structure selected" (14). Similarly, Grennan and Jablonski (1999) believe that student affairs professionals "need more understanding of the skills necessary for conceptualizing and conducting research as well as the types of research questions that would be valuable in improving programs and services" (80).

To that end, the "Principles of Good Student Affairs Practice" (NASPA/ACPA, 1997) cites using assessment methods to gain "high-quality" information about our students' experiences in order to improve student and institutional performance. Relationships forged across departments and divisions – in this case between student affairs and Institutional Research – affirm shared educational goals for our students' success. Moxley (1999) believes that establishing a relationship with "information-rich administrative offices," such as institutional research, can be critical to meeting student affairs data needs (14). Institutional research units often produce periodic reports and research findings useful for shaping student affairs goals and objectives. Institutional research staff can provide technical expertise in selecting research samples, determining data collection methodologies, and refining instruments. Being aware of the development of a comprehensive survey instrument, for instance, and the ability to add questions illustrates the benefits of a strong communication link with the institutional research office (Moxley, 1999, 15).

Similarly, when it comes to undertaking comprehensive research efforts in student affairs-related areas, higher priorities may prevail and the concept of somehow measuring student development, student services and a students' co-curricular experience may not be within the institutional researchers' areas of expertise. A strong relationship, in the

form of on-going and deliberate collaboration, is needed to “ensure that this information-fertile area is maximally used by student affairs” and the institution more broadly (Moxley, 1999, 14-15).

**Data Sources**

Since 2007 the Division of Student Success and the Office of Institutional Research, Planning and Effectiveness at the University at Albany have more systematically partnered on several dozen studies of students’ experience. These studies have included national benchmarks (NSSE and the Profile of Today’s College Student), institutional and system-wide studies (the State University of New York’s Student Opinion Survey) as well “home grown” assessments of student satisfaction, learning through Residential Life (ACUHO-I/EBI), Campus Center Management (ACUI), Orientation (student and parent’s experience), Fraternity and Sorority Affairs and the Disability Resource Center (student and faculty perceptions) as well as post-graduation plans through Career Services.

Appendix A provides a sample of student affairs-related assessment activities since 2000. The table highlights institutional assessments, facilitated by institutional research, as well as unit-level assessments administered by student affairs areas to supplement the institutional findings. In each instance listed, units were informed by findings from institutional assessments and sought to examine broad issues more specifically with unit-level analyses. While there is a story to each assessment and question included in Appendix A, we have chosen to address two “cases” of student

affairs units using findings from an institutional assessment to adopt an instrument to explicitly assess unit-level programs, activities and services.

The SUNY Student Opinion Survey (SOS) has been given to undergraduate students at all of SUNY's colleges and universities every third year since 1985. This survey helps UAlbany assess various areas of the academic experience, highlighting the areas where we are doing well, but more importantly, identifying the areas where we need to improve. The most recent administrations of the SOS were in the Spring Semesters of 2000, 2003, 2006 and 2009. The SOS is UAlbany's most important general survey of student satisfaction, in that it asks our undergraduates about their experiences and satisfaction with a wide variety of aspects of university life, including academic and non-academic facilities and experiences. As Appendix A shows, the SOS will typically ask anywhere from one to a handful of questions on a particular area, making it valuable for use as a broad gauge of student satisfaction in a large number of areas, but less valuable for getting into the details of what works and what does not in those areas. In order to determine what concrete steps can be taken to improve student experiences and satisfaction, it is necessary to conduct topical assessment surveys that delve more deeply in a particular area of student life.

An example of this sort of detailed topical survey is the Association of College and University Housing Officers International (ACUHO-I) resident student assessment, administered five times since 2001. In prior administrations, this has been a paper instrument that residential life staff members deliver to individual student rooms; this Fall, however, the survey is being administered via the internet. The instrument has

traditionally been administered in November to a sample of 3,000 resident students. Response rates with the paper administration have varied between 72% and 89%.

Another example is Orientation's summer planning conference evaluations, administered after each orientation program to incoming students and their families, is a paper instrument included in participant's orientation packet and collected at the conclusion of the program. The evaluation is administered to the population of incoming students, between 2,800 and 3,400 each year with a response rate ranging between 96% and 98% annually. The 2008 National Orientation Directors Association (NODA) benchmarking instrument was administered to all incoming students who participated as part of that summer's orientation program (3,358 students' total) of which 36% responded (1,222 students).

**Case #1: Residential Life – Developing a Learning Outcomes Programming Model**

In 2000, the SUNY-wide student opinion survey (SOS) indicated that UAlbany's students were very dissatisfied with their experience with our residence halls. Only 19% of students surveyed indicated that they were satisfied or very satisfied with "residence hall services and programs," with 46% expressing dissatisfaction and the remaining 35% neutral. Opinions about the "general condition of residence hall facilities" were even worse – 18% were satisfied and 58% were dissatisfied and 24% neutral. These figures were even lower than the previous two SOS administrations in 1994 and 1997.

Because of these poor results, the following year Residential Life began participating in ACUHO-I's resident satisfaction survey, administered nationally, to gauge how their programs and services measured against peer institutions. Since then,

Residential Life has utilized that instrument four additional times. The department has experienced noticeable improvements in students' overall satisfaction with their residential experience. Satisfaction with residence hall services and programs increased from 19% in 2000 to 27% in 2003 and 38% in 2006, as measured by the SUNY Student Opinion Survey.

One area in particular that had routinely been rated low when compared to peer institutions was the delivery of programs in the residence halls. As a result, in 2006 the department undertook a comprehensive overhaul of its programming model with an emphasis on student learning.

Past programming models at the University had focused on "categories" of activities, and success was typically defined in terms of attendance and advertising. The new model changes the paradigm by defining success in terms of the evidence presented. All educational programming in the department must address one or more of these overarching learning outcomes.

This model incorporates two levels of outcomes assessment. The first uses the within-program learning piece to show the program's worth as described in the preceding sections. The second or "macro-level" assessment comes from the analysis of the department's biannual ACUHO-I survey of the residence halls and apartments. It is here that the outcomes of the program are truly noticeable.

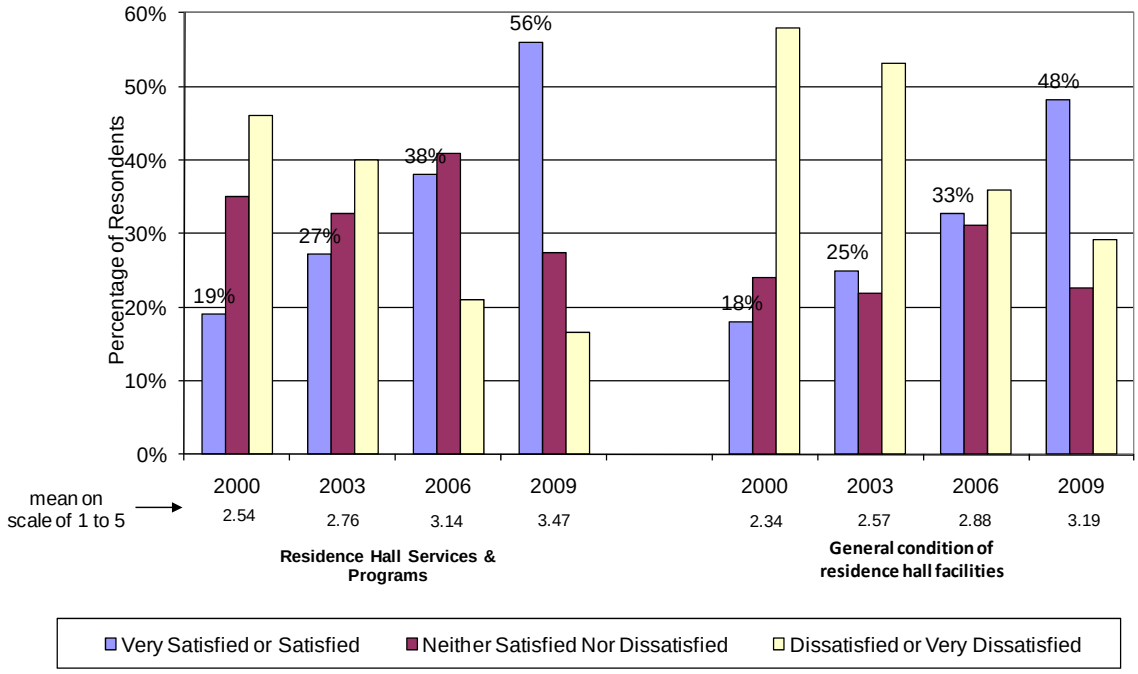
Since the model has been in use, the ACUHO-I survey has only been administered once but the results are telling. Compared to students who did not attend residential life programming, students who attended programming in the residence halls had higher satisfaction with certain factors such as managing time, studying, and solving

problems as well as personal interactions. In these factors, UAlbany students were more satisfied than its peers in the six selected comparison institutions, its peers in its Carnegie Class and, in the case of managing time, than all surveyed institutions.

The Learning Outcomes Showcase is the culmination of each semester's programs and displays the programs which accomplished their outcomes and showcases the student learning that occurred. A presentation board is created for each award recipient (there are usually three – gold, silver, bronze for each overarching outcome) and the student learning – original work, video, evaluative tool, etc. – is on display. The University community is invited to the awards ceremony but the presentation boards are left on display for passers-by to see the great work Residential Life does on a daily basis.

Overall, the programming model has been successful in terms of the number of quality of programs produced by department staff as well as in terms of the satisfaction of students with their residential experience. By the time of the 2009 administration of the SOS, satisfaction with residence hall services and programs had climbed to 56% -- up all the way from 19% in 2000, as shown in Figure 2, below.

Figure 2. Improvement in Student Satisfaction with Residence Halls, 2000-2009, as Measured by SUNY Student Opinion Surveys




**Case #2: Orientation – Enhancing Transfer Students' Orientation Experience**

In 2000, the SUNY-wide student opinion survey (SOS) indicated levels of satisfaction with "new student orientation services" that, while substantially higher than the residence halls, still indicated room for improvement. Among the students who participated in the 2000 SOS, 20% expressed dissatisfaction with orientation, but only 47% were satisfied (with the remaining 33% neutral). Those figures improved somewhat in 2003 (15% dissatisfied, 51% satisfied, 34% neutral) and 2006 (14% dissatisfied, 55% satisfied, 31% neutral).

The University's orientation office has been administering program evaluations since the early 1980s to all students participating as part of their summer planning

conferences. In that time, the office had never participated in a benchmarking study to evaluate how its program compared to peer institutions. In 2008, orientation participated in NODA's benchmarking survey. As a standalone evaluation of the University's orientation program, the instrument did not reflect program deficiencies. Students largely seemed satisfied with their orientation experience. However, upon benchmarking transfer students' orientation experience with those transfer students at peer institutions, orientation staff discovered that transfer students at the University at Albany were less satisfied than those students at peer institutions. Further examination of program-level evaluations indicated similar findings.

As a result of Orientation's 2008 benchmarking survey and summer program evaluations, the transfer student program was targeted for improvement. These instruments pointed out that transfer students desired: 1) more interaction with staff, 2) personal interaction with one another, and 3) a greater sense of connection to the UAlbany community.

In response, Orientation modified the time allocated for the Resource Fair so that all students and family members could attend and added additional offices and departments to allow for the convenience of interactions. Additionally, Orientation allocated time for an interactive session with Orientation Assistants in small groups to allow students an opportunity to get to know one another. Orientation also collaborated with Transfer Experience Coordinator to enhance the program content of the Conference to include discussions about resources available on campus specifically targeting a wide variety of transfer students needs (i.e.: Mentoring Program, Tau Sigma, Transfer Resource Guide and the Driving Force). Perhaps as a result, satisfaction with new student

orientation services improved substantially in the 2009 SOS – 13% dissatisfied, 64% satisfied, and 23% neutral.

**Conclusion**

The assessment of student's co-curricular experiences, including a variety of student services, programs and activities, requires thoughtful collaboration between student affairs units and institutional research offices. Partnering to ensure deliberate assessment of students' co-curricular experiences benefits student affairs' by providing findings that seek to improve programs, services and activities while providing institutional research offices with data that compliment the findings collected from students' experiences while at college.

The example of student affairs-institutional research collaboration at the University at Albany lends itself to at least three "lessons" for what a health collaborative relationship look like: time, people and reciprocity.

**Lesson #1: "It Takes Time"**

While the University's institutional research office had, for well over a decade, supported student affairs' assessment efforts, it has only been since 2007 that units throughout student affairs fully began to embrace assessment at the unit level. As outlined in Table 1, units' supplemental assessment efforts gleaned greater clarity for purposes improving programs and services for our students.

It has been student affairs' approach since 2007 that "good assessment takes time" and that working slowly and systematically towards a comprehensive assessment agenda for the division will benefit its units, staff and students the most in the long-run. Similarly, institutional research has worked to build relationships with student affairs professionals so as to leverage their collective energies towards institutional assessment and planning efforts.

Collaborative efforts built on trust, mutual understanding and the shared goal of finding good data to inform decision-making which ultimately benefits a University's students is constantly being built upon over time.

**Lesson #2: "Many Hands Make Light Work"**

Whereas nearly a decade ago a single staff person, in institutional research, provided "expertise" in the area of assessment – statistical analysis, summarizing data and providing recommendations for improvement – today, nearly 30 staff provide leadership for assessment in student affairs-related units.

While the staff in institutional research providing support to student affairs "doubled" since 2005 (from one to two), the number of staff in student affairs charged

with assessment grew exponentially. The establishment of an assessment position in the student affairs central office, as well as the designation of assessment coordinators in each of student affairs' 13 units has contributed to a noticeable increase in the number of assessment activities across student affairs units.

Furthermore, building upon broad-based, institution-wide assessments with unit-specific program evaluations and the assessment of services empowers student affairs staff to "do more" with findings to enhance students' experience at the University. It is not unusual for student affairs professional to propose a series of questions to be included as part of an institutional-wide assessment administered through or in partnership with institutional research. Similarly, institutional research will routinely reach out to student affairs colleagues to include questions or encourage the assessment of specific activities to support an institution-wide assessment.

**Lesson #3: "True Collaboration is a Team Effort"**

For the better part of the last two decades, institutional research at the University at Albany providing leadership and guidance on all student affairs-related assessment. Assessment findings were shared with key constituencies and decision-makers. Similarly, when select units (i.e.: Residential Life, University Counseling Center) began to administer assessments of their own, their findings were also shared with institutional research. The challenge historically had been actually doing something with the assessment findings. With limited understanding and comfort levels among student affairs staff with respect to good assessment practices, staff was often hesitant to fully immerse themselves and their units into assessment findings.

Today, collaboration between institutional research and student affairs is a cornerstone of comprehensive, thoughtful assessment practices. Annual assessment schedules are created the summer prior to the start of the academic year and are shared with institutional research for institution-wide planning purposes. Findings are summarized, publicized and prioritized. Program enhancements are detailed as part of individual units' annual reports. All of which are shared with institutional research in real time. Student affairs professionals charged with assessment in their area are not only familiar with the important work of colleagues in institutional research, but they routinely reach out to these colleagues for guidance and insights. Additionally, professionals in institutional research welcome the opportunity to provide feedback and analysis of findings provided to them by colleagues in student affairs.

Student affairs' annual day-long assessment retreat – held at the beginning of June – expressly includes colleagues from institutional research who celebrate the collective successes of student affairs units' assessment efforts and program improvements.

Finally, the results of this collaboration received very favorable commentary in the University's 2010 Middle States review team report, stating that "Student Services (Student Success) assessment activities are very robust, with a five-year history…These assessment tools and the information they collect are used to improve programs and services."

***Appendix A:***

*Sample of Institutional and Unit-level Student Affairs-related assessment activities (2000 – 2010)*

| Topic/ Unit | Institutional Assessment SUNY Student Opinion Survey (SOS) Question(s) | Unit-Level Assessment(s) | Question(s) |
|---|---|---|---|
| **Residence Halls** | **(Satisfaction with)**<br>**3a11:** General Condition of Residence halls<br>**3a40:** Residence hall Services/ Programs<br>**3a41.** Clarity of residence hall rules/policies | ACUHO-I<br><br>(2001, 2002, 2004, 2006, 2008) | **(Satisfaction with)**<br>**Q33-Q40**. Facilities.<br>**Q26-Q29.** Providing various programming.<br>**Q034 – Q047**. Providing various services.<br>**Q48-Q50**. Room Assignment or Change Process. |
| **Campus Center** | **(Satisfaction with)**<br><br>**3a10:** Campus Center/Student Union | Campus Center Survey (ACUI, 2008) | **Q23 – Q31.** Campus Center as a facility that…[student perceptions].<br>**Q33**. Promotes a sense of community on campus.<br>**Q35**. Is an enjoyable place to spend time.<br>**Q36**. Is a place where I feel welcome.<br>**Q38**. Is a student-oriented facility. |
| **Campus Safety** | **(Satisfaction with)**<br><br>**3a14:** Personal Safety/Security on this campus | Campus Safety Survey (ASCA, 2009) & Annual Safety Survey (handheld devises, 2009 & 2010) | **Q01-Q17**. How safe do you feel [various times, locations]..?<br>**Q20**. How often, on average, do you see campus safety officers patrolling the campus?<br>**Q28.** Campus security/campus police are responsive to campus safety issues.<br><br>**Q01**. How safe do you feel on campus overall.<br>**Q03**. Adequate campus safety/campus police presence on campus |
| **Health & Wellness** | **(Satisfaction with)**<br>**3a23:** Educational Programs regarding alcohol and substance abuse<br>**3a24.** Sexual assault prevention programs<br>**3a26.** Student health programs<br>**3a39.** Personal counseling services | Health Center Survey(s) (ACHA, 2009) & Semester user evaluation (2009, 2010) | **Q22**. The provider listened carefully to your concerns.<br>**Q24**. Quality of the explanations and advice you were given by your provider for your condition and the recommended treatment.<br>**Q25**. Quality of the explanations and advice you were given for your condition and the recommended treatment. |

| | | | |
|---|---|---|---|
| **Career Planning/ Job Placement** | **(Satisfaction with)**<br><br>**3a29.** Career Planning Services<br>**3a30.** Job Placement Services | Student Experience Survey (2005, 2007) & Survey of Recent Graduates (perennial) & Counselor Feedback (2009 & 2010) | **Q8**. The counselor was knowledgeable about the topics that we discussed.<br>**Q9**. The counselor encouraged me to think more about career-related issues.<br>**Q10**. I learned more about the topic/s I chose above than I previously knew. |
| **Student Activities** | **(Satisfaction with)**<br><br>**3a31.** Purposes for which student activity fees are used<br>**3a33.** College social activities<br>**3a38.** Opportunities for involvement in campus clubs/activities<br>**3a42.** Student newspaper<br>**3a43.** Student radio station | Student Activity Survey (NACA, 2009)<br><br>& Campus Recreation Survey (NIRSA, 2010) | **Q4**. To what degree are there student activities on your campus that interest you?<br>**Q10**. Generally, how involved are you in campus activities at this college/university?<br>**Q25**. Are you as involved in campus activities as you would like to be?<br>**Q30**. As a result of participating in campus activities... - I have been able to connect with other students.<br><br>**Q79**. Number of team intramural sports offered.<br>**Q81**. Number of Club Sports offered |
| **Orientation** | **(Satisfaction with)**<br><br>**3a28.** New student orientation | Orientation program evaluations (perennial)<br><br>& Orientation benchmark (NODA, 2008) | **Q3**. The orientation leaders and staff helped me feel welcome at UAlbany.<br>**Q5**. I learned things that will help ease my transition to UAlbany.<br><br>**Q8-Q10**. Orientation staff.<br>**Q11**. Orientation helped me to know what to expect academically at UAlbany.<br>**Q12**. Orientation helped me to know what to expect socially at UAlbany.<br>**Q13**. I met new people at orientation that I am still friends with. |
| **Judicial Affairs** | **(Agreement with)**<br><br>**3b2.** The rules governing student conduct are clear to me | Judicial Affairs survey (ASCA, 2009)<br>& Referred Party follow-up survey. | **Q70 (Q5)**. Hearing/judicial process, I was treated fairly.<br>**Q72 (Q7)**. Hearing/judicial process, all of my questions were answered.<br>**Q74 (Q9)**. I believe that the sanctions that I was assigned were educational in nature. |

## REFERENCES


American Association for Higher Education, American College Personnel Association, & National Association of Student Personnel Administrators (AAHE, ACPA, & NASPA). (1998). *Powerful partnerships: A shared responsibility for learning*. Washington, DC: ACPA.

American College Personnel Association (ACPA). (1994). *The student-learning imperative: Implications for student affairs*. Washington, DC: ACPA.

Grennan, K. F. and Jablonski, M. A. (1999). "Collaboration Between Student Affairs and Faculty on Student-Related Research." *New Directions for Student Services*, No. 85, Spring: Jossey-Bass Publishers.

Hyman, R. (1995). "Creating campus partnerships for student success." *College & University*, 71(2), 2–8.

Kezar, A. (2003). "Enhancing Innovative Partnerships: Creating a Change Model for Academic and Student Affairs Collaboration." *Innovative Higher Education*, Vol. 28, No. 2, Winter.

Knefelkamp, L. L. (1991). *The seamless curriculum. CIC Deans Institute: Is this good for our students?* Washington, DC: Council for Independent Colleges.

Kuh, G. D., Douglas, K. B., Lund, J. P., & Ramin Gyurmek, J. (1994). *Student learning outside the classroom: Transcending artificial boundaries*. ASHE-ERIC Higher Education Report, No. 8.Washington, DC: The GeorgeWashington University, Graduate School of Education and Human Development.

Kuh, G. D. (1996). Guiding principles for creating seamless learning environments for undergraduates. *Journal of College Student Development, 37,* 135–148.

Kuh, G. D. and Banta, T. W. (2000). "Faculty–Student Affairs Collaboration on Assessment: Lessons from the Field." *About Campus*. January/February.

Martin, J., & Murphy, S. (2000). *Building a better bridge: creating effective partnerships between academic and student affairs*. Washington, DC: National Association of Student Personnel Administrators.

Moxley, L. S. (1999). "Student Affairs Research and Evaluation: An Inside View." *New Directions for Student Services*. No. 85, Spring: Jossey-Bass Publishers.

National Association of Student Personnel Administrators (NASPA) (1997). *Principles of good practice for student affairs*. Washington, DC: NASPA.

National Association of Student Personnel Administrators (NASPA) (2009). *Assessment Education Framework*. Washington, DC: NASPA.

Schroeder, C. S., & Hurst, J. C. (1996). Designing learning environments that integrate curricular and cocurricular experiences. *Journal of College Student Development*, *37*, 174–181.

UAlbany 1. The Albany Outcomes Assesment Model home page. http://www.albany.edu/assessment/ualb_outcomes_model.html.

UAlbany 2. The Albany Outcomes Assessment Model, Brief. http://www.albany.edu/assessment/albany_model_brief.doc.

# DIRECT VS. INDIRECT MEASURES OF STUDENT LEARNING: A COMPARISON OF MATCHED RESULTS FROM THE COLLEGIATE LEARNING ASSESSMENT AND THE NATIONAL SURVEY OF STUDENT ENGAGEMENT

**Dr. Braden J. Hosch**
**Director of Policy, Finance, and Academic Affairs**
**Connecticut Department of Higher Education**
**bhosch@ctdhe.org**

**********

This paper examines matched unit-record results of 217 students who took both the Collegiate Learning Assessment (CLA) and the National Survey of Student Engagement (NSSE) over a three-year period at a public Master's-Larger Programs institution in the Northeast. Results indicate that seniors recruited to take CLA using incentives exhibited more engaged behaviors on a range of NSSE items compared to seniors who did not take CLA, suggesting a recruitment bias in the testing population. Further, results confirm previous research by Carini, Kuh, and Klein (2006) that indicates only small relationships between test scores and survey items.

**********

Recent accountability initiatives in higher education have called for the direct assessment of student learning in ways that provide comparable information across institutions and states (Commission on the Future of Higher Education, 2006; Miller, 2006). Of particular note, the Voluntary System of Accountability (VSA) prompts public institutions to administer common standardized instruments to measure student learning and to examine value added by institutions to the educational experience (McPherson & Shulenburger, 2006).

The VSA requires participating institutions to administer one of three standardized instruments to measure student learning and to demonstrate the value-added to learning by the institution. A recent validity study conducted by the test owners indicates the tests are valid and reliable (Klein, Liu, & Sconing, 2009), but it is important to contextualize these claims to recognize they mean that the same students under the same testing conditions will perform about the same way on any one of the three instruments.

VSA requires participating institutions to administer one of three standardized instruments to measure student learning and to demonstrate the value-added to learning by the institution. These three instruments are the Collegiate Assessment of Academic Proficiency (CAAP) owned by ACT, Inc., the Measure of Academic Proficiency and Progress (MAPP) owned by the Educational Testing Service, and the CLA owned by the Council for Aid to Education.

CLA's measurement construct for evaluating the value added by institutions adopts a cross-sectional design with institutions administering tests to samples of at least 100-200 first-year students and 100-200 graduating seniors who began their undergraduate experience at the institution. Through 2008-09, scores on the tests were compared to an expected score based on SAT or ACT scores, and a relative-to-expected score was calculated as the residual between the actual and expected scores (performance categories were then described as "well above expected," "above expected," "at expected," "below expected," and "well below expected"). Further an institutional value-added score is calculated by subtracting the first-year residual from the senior residual (Klein, et al., 2007; Steedle, 2009). For instance, if entering first-year students score at expected levels while seniors score well above expected the institution's value-added

score will likely also be above or well above expected. Conversely, for institutions at which first-year students score above expected levels but seniors score at expected levels, the institutional value added might be below expected, depending on the magnitude of the score differential.

CLA and the VSA have been criticized for use of a cross-sectional methodology to established educational value-added (Garcia, 2007; Banta & Pike, 2007; Kuh, 2006). Borden and Young (2008) provide an eloquent and comprehensive examination of the deployment of validity as a construct, using CLA and the VSA as a case study, to highlight the contextual and contested nature of validity across various communities. Further, student motivation, the amount of time spent on the test, and test administration procedures appear to be related to test performance, suggesting that direct measures of student learning may not yet be nuanced enough to anchor accountability systems (Steedle, 2010; Hosch 2010).

Testing organizations have tried to answer these charges (Klein, Benjamin, Shavelson, & Bolus, 2007), perhaps most effectively by demonstrating the utility of their instruments in longitudinal administrations to the same students (Arum & Roksa, 2008), although such practices can be prohibitively expensive and take years to produce results.

To provide some measure of validity to CLA and also to NSSE, Carini, Kuh, and Klein (2006) examined the relationship between results from the NSSE and CLA of 940 students from 14 institutions in 2002, and valuably determined that many indirect (self-reported) measures of student learning on the NSSE were positively associated with performance on the CLA, although most of these relationships were weak in strength. While this study is useful in providing some cross-validation of each of these instruments, it has some limitations. Because 48% of the students in the study population were sophomore and juniors, and because the 940 students were distributed across 14 institutions, each institution contributed on average only about 19 freshmen and 15 seniors (actual numbers may have varied). Further, 10 out of the 14 participating institutions (71%) were private, while 61% of students attend public institutions (National Center for Educational Statistics, 2009).

The current study answers a call from Carini, Kuh, and Klein to examine the relationships between the direct measures of student learning generated by CLA with indirect measures of student learning obtained on NSSE. The linkage of CLA results with NSSE results answers two basic questions about the relationships between the two instruments.

First, do students who took the CLA constitute a representative sample of students at the institution? While typical participant characteristics like gender, race, field of study, SAT scores, and grade point average can provide some insight to this first question, these indicators do not constitute the rich data

The study also provides some lessons about the utility and practicality of linking results from these different measures of student learning. Finally, the results may provide evidence for or against the challenge levied by Porter (2009) that many widely used instruments in higher education do not meet generally accepted standards of validity.

**Methodology**

The present study includes only about half the number of first-year students and seniors in the Carini, Kuh, and Klein (2006) project but limits the population to a single institution, a public Master's-Larger Programs institution in the northeast. CLA and NSSE were administered during the 2007-08, 2008-09, and 2009-10 academic years. Across all years a total of 972 first-year students and 1,004 seniors participated in NSSE and 345 first-year students and 338 seniors

participated in NSSE. A total of 93 first-year students and 124 seniors completed both the NSSE and the CLA.

*CLA Recruitment and Administration*

For the present study, the CLA was administered to first-year students and seniors in 2007-08, 2008-09, and 2009-10 with the ultimate intention of publishing the scores on the institution's VSA College Portrait. CLA is a 60- to 90-minute constructed response assessment that is administered online under the supervision of a local proctor. Student recruitment posed difficulties in all test administrations because many identified, eligible participants balked at the prospect of taking a 90-minute essay test. Thus, while 683 students were tested over three years, obtaining a representative sample of students to take CLA constituted an ongoing challenge in test administration.

First-year students were recruited by targeting selected sections of the institution's first-year experience course. Students were more or less randomly assigned to these sections, and instructors of these courses used a variety of methods to encourage participation. First-year students who took CLA were more or less representative of the student body as a whole in terms of race/ethnicity, gender, SAT scores, high school class rank, and field of study. It is valuable to observe, however, that while the sample was representative, it was by no means random.

Recruitment of seniors posed more significant challenges, and recruitment practices evolved of the course of the first test administration. In the 2007-08 administration, about half of the participants came from three senior capstone courses (psychology, social work, and management), while the remaining half were recruited by means of a $25 discount on graduation regalia. Subsequent administrations in 2008-09 and 2009-10 did not involve senior capstone courses and instead recruited graduation seniors to participate by offering a full waiver of regalia fees ($35 in 2009 and $40 in 2010). Again, these procedures did not yield random samples, but in Spring 2009 and Spring 2010, students participating were roughly representative of the graduating class, with 41-45 majors represented in each term (compared to just 24 different majors in Spring 2008) and clusters of 10-12 students in areas in which students earn the highest portion of degrees (business, education, and psychology). See Hosch (2010) for additional details about CLA administration and some of the limitations of these recruitment methods.

*NSSE Recruitment and Administration*

Administration of NSSE on this campus was conducted solely online in the spring semester of each year to capture information from first-year students and graduating seniors. No efforts were made specifically to recruit students who also participated in CLA, an unlike the study design of Carini, Kuh, and Klein where the 25 minute NSSE was administered following CLA, the NSSE was administered separately from CLA, in many instances months later.

Local administration of NSSE was subject to several strictures and controls from NSSE that limited direct contact with potential participants. Potential participants were contacted directly only five times via email, and signs were posted around campus encouraging students to participate, but no other direct contact was allowed. Incentive for participation in each of these years was entry into a random drawing for one of two iPhones.

Response rates for the three years ranged between 22% in spring 2008 to 29-30% in spring 2009 and spring 2010. Populations responding to NSSE were generally representative of students at the institution in terms of race/ethnicity and field of study. However, women were overrepresented by about 10%, and NSSE respondents in general registered an average cumulative GPA about 0.2 grade points higher than the student population at-large. Such

population differences are not unusual for participation in NSSE or other higher education surveys, although they may have implications for how to interpret results (Clarkberg, Robertson & Einarson, 2008).

Table 1. NSSE and CLA Participants by Year

| | Took NSSE | 2007-08 | | | 2008-09 | | | 2009-10 | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Took CLA | | | Took CLA | | | Took CLA | | | Took CLA | | |
| | | Yes | No | Total | Yes | No | Total | Yes | No | Total | Yes | No | Total |
| First-Year Students | Yes | 27 | 270 | 297 | 31 | 326 | 357 | 35 | 284 | 319 | 93 | 880 | 973 |
| | No | 78 | 0 | 78 | 79 | 0 | 79 | 95 | 0 | 95 | 252 | 0 | 252 |
| | Total | 105 | 270 | 375 | 110 | 326 | 436 | 130 | 284 | 414 | 345 | 880 | 1225 |
| Seniors | Yes | 34 | 227 | 261 | 54 | 359 | 413 | 36 | 357 | 393 | 124 | 943 | 1067 |
| | No | 65 | 0 | 65 | 80 | 0 | 80 | 69 | 0 | 69 | 214 | 0 | 214 |
| | Total | 99 | 227 | 326 | 134 | 359 | 493 | 105 | 357 | 462 | 338 | 943 | 1281 |
| All Students | Yes | 61 | 497 | 558 | 85 | 685 | 770 | 71 | 641 | 712 | 217 | 1823 | 2040 |
| | No | 143 | 0 | 143 | 159 | 0 | 159 | 164 | 0 | 164 | 466 | 0 | 466 |
| | Total | 204 | 497 | 701 | 244 | 685 | 929 | 235 | 641 | 876 | 683 | 1823 | 2506 |

The number of cases in which students took both CLA and NSSE in the same year was low. Just 93 first-year students took CLA in a fall semester and then NSSE in the subsequent spring over the course of three years. Similarly, just 124 seniors took CLA and NSSE in the same spring, for a total of 217 students in the sample. These numbers still dwarf the estimated same institution samples obtained by Carini, Kuh, and Klein by a multiple of 5 to 10, but they still represent not even 9% of the students who took NSSE over this period, and even this group represented just over 25% of students invited to take the NSSE. The bottom line is that caution should be used before generalizing these results to the institutional level or beyond.

## Findings

Overall findings suggest that students exhibited levels of engagement just below the 50$^{th}$ percentile on NSSE benchmarks (Hosch & Joslyn, 2010) and students performed in 37$^{th}$-70$^{th}$ percentiles on CLA, depending on the semester. Entering academic ability (as measured by SAT scores) and the amount of time students spent on the test were the factors most related to CLA performance (Hosch, 2010). An attenuated summary of results is provided here for reference.

First-year students who took CLA were more or less representative of students who completed NSSE at this institution, but seniors who took CLA reported higher levels of engagement on multiple survey items. This analysis was conducted in a fashion quite similar to the NSSE Means and Frequencies report for institutions which conducts t-tests between groups and expresses significant differences in terms of standard deviations or effect size.

Table 2. NSSE Benchmarks and CLA Performance

All figures represent percentiles for the entire test/survey universe unless noted

| | 2007-08 | 2008-09 | 2009-10 |
|---|---|---|---|
| First-Year Students | | | |
| NSSE Benchmarks* | | | |
| Level of Academic Challenge | 46 | 44 | 46 |
| Active and Collaborative Learning | 44 | 42 | 45 |
| Student-Faculty Interaction | 50 | 50 | 52 |
| Enriching Educational Experiences | 48 | 43 | 40 |

| | 2007-08 | 2008-09 | 2009-10 |
|---|---|---|---|
| Supportive Campus Environment | 43 | 47 | 50 |
| | | | |
| CLA Scores | | | |
| Raw score | 51 | 67 | 53 |
| Adjusted score | 62 | 84 | --† |
| Relative-to-Expected Performance | At | Above | --† |
| Minutes spent | -- | 49 | 44 |
| | | | |
| Senior | | | |
| NSSE Benchmarks* | | | |
| Level of Academic Challenge | 44 | 45 | 47 |
| Active and Collaborative Learning | 45 | 48 | 49 |
| Student-Faculty Interaction | 47 | 48 | 47 |
| Enriching Educational Experiences | 44 | 44 | 44 |
| Supportive Campus Environment | 40 | 44 | 44 |
| | | | |
| CLA Scores | | | |
| Raw percentile score | 37 | 70 | 62 |
| Adjusted percentile score | 63 | 98 | --† |
| Relative-to-Expected Performance | At | Well Above | --† |
| Minutes spent | 45 | 63 | 55 |
| | | | |
| CLA Institutional Metrics | | | |
| Adjusted Percentile for "Value Added" | 49 | 79 | 74 |
| Performance Relative to Other Institutions | At | Above | Near† |

* Percentiles for NSSE benchmarks calculated from effect sizes (z-scores) and assume a normal distribution; some estimations here will exceed other representations of institutional performance by 5-7 percentile points. NSSE sensibly halted the practice of calculating institutional percentiles in 2007 because variation within institutions substantially outstrips variation among institutions.

†Beginning in 2009-10 CLA replaced ordinary least squares (OLS) with hierarchical linear modeling (HLM) to determine institution value-added scores. Use of HLM helps to control for nested effects of institutions, but it also does not provide an adjusted CLA score for groups of students. Also, the phrase "at expected" performance was replace with "near expected."

For first-year students, only five NSSE items exhibited significant differences, including reporting higher levels of growth in contributing to the welfare of their community, talking with faculty members or advisors about career plans, and participating in service learning. They were also less likely to have serious conversations with students who were different from them and were more likely to spend more time watching television and relaxing than first-year students who did not take the CLA. In all of these instances, the differences were at a level generally deemed small (effect size, or Cohen's *d,* between 0.20 and 0.29)

On the other hand, seniors who took the CLA exhibited significant differences on 23 NSSE compared to seniors who did not take CLA. Seniors who took the CLA on average worked fewer hours off-campus for pay ($d$ = -0.48) and more often worked on a research project with a faculty member on a research project ($d$=0.42) or on other activities outside of coursework ($d$=0.41 than did seniors who did not take CLA. A range of other differences included CLA takers spending more time on community service, working for pay on-campus, participating in activities to enhance their spirituality, and a range of other behaviors that typically are associated with deeper engagement with the undergraduate educational experience.

Table 3. Differences on NSSE items between FIRST-YEAR students who did and did not take CLA

| NSSE Item | Did not take CLA n=880 | Took CLA n=93 | Sig | Effect Size |
|---|---|---|---|---|
| Institutional contribution: Contributing to the welfare of your community | 2.26 | 2.53 | * | 0.29 |
| Talked about career plans with a faculty member or advisor | 2.18 | 2.41 | * | 0.26 |
| Had serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values | 2.64 | 2.43 | * | -0.24 |
| Hours per 7-day week spent relaxing and socializing (watching TV, partying, etc.) | 4.01 | 4.39 | * | 0.23 |
| Participated in a community-based project (e.g., service learning) as part of a regular course | 1.52 | 1.70 | * | 0.21 |

Table 4. Differences on NSSE items between SENIORS who did and did not take CLA

| NSSE Item | Did not take CLA n=943 | Took CLA n=124 | Sig | Effect Size |
|---|---|---|---|---|
| Hours per 7-day week spent working for pay OFF CAMPUS | 5.15 | 3.87 | *** | -0.48 |
| Work on a research project with a faculty member outside of course or program requirements | 2.18 | 2.59 | *** | 0.42 |
| Worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.) | 1.72 | 2.13 | *** | 0.41 |
| Community service or volunteer work | 3.00 | 3.39 | *** | 0.37 |
| Talked about career plans with a faculty member or advisor | 2.42 | 2.76 | *** | 0.36 |
| Hours per 7-day week spent working for pay ON CAMPUS | 1.46 | 1.98 | *** | 0.35 |
| Exercised or participated in physical fitness activities | 2.52 | 2.84 | *** | 0.31 |
| Participated in activities to enhance your spirituality | 1.74 | 2.06 | *** | 0.31 |
| Tutored or taught other students (paid or voluntary) | 1.81 | 2.12 | ** | 0.30 |
| Practicum, internship, field experience, co-op experience, or clinical assignment | 3.19 | 3.47 | *** | 0.30 |
| Hours per 7-day week spent participating in co-curricular activities (organizations, campus publications, student government, fraternity or sorority, intercollegiate or intramural sports, etc.) | 1.78 | 2.21 | ** | 0.29 |
| Participated in a community-based project (e.g., service learning) as part of a regular course | 1.60 | 1.87 | ** | 0.28 |
| Had serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values | 2.66 | 2.93 | ** | 0.28 |
| Participate in a learning community or some other formal program where groups of students take two or more classes together | 2.42 | 2.67 | * | 0.25 |
| Culminating senior experience (capstone course, senior project or thesis, comprehensive exam, etc.) | 2.72 | 2.98 | ** | 0.25 |
| Hours per 7-day week spent providing care for dependents living with you (parents, children, spouse, etc.) | 2.65 | 2.16 | ** | -0.25 |
| Put together ideas or concepts from different courses when completing assignments or during class discussions | 2.90 | 3.07 | * | 0.22 |
| Tried to better understand someone else's views by imagining how an issue looks from his or her perspective | 2.78 | 2.97 | * | 0.22 |
| Institutional contribution: Working effectively with others | 3.10 | 3.28 | * | 0.22 |
| Discussed ideas from your readings or classes with faculty members outside of class | 2.12 | 2.33 | * | 0.21 |
| Attended an art exhibit, play, dance, music, theater, or other performance | 1.99 | 2.18 | * | 0.21 |
| Used an electronic medium (listserv, chat group, Internet, instant messaging, etc.) to discuss or complete an assignment | 2.72 | 2.91 | * | 0.20 |
| Had serious conversations with students of a different race or ethnicity than your own | 2.69 | 2.89 | * | 0.20 |

Differences between seniors who took the CLA and those who did not take CLA likely indicate self-selection bias in the population that took the test, and so the NSSE results provide strong evidence that the CLA-taker did not constitute a representative sample in terms of behavior and engagement in the life of the university. This phenomenon was less observable among the first-year students who took both CLA and NSSE perhaps because of embedding the recruitment practices in first-year experience classes and perhaps also because first-year students may simply be more tractable than seniors.

In terms of the relationship between performance on CLA and NSSE items, only limited correlations were observed. In large part, this finding supports the general conclusion that was reach by Carini, Kuh, and Klein that the relationships between engagement and outcomes are relatively small in magnitude. The present study actually found slightly stronger relationships at the item level, and no relationships with the engagement benchmark scales.

Table 5. Correlations between NSSE Items and CLA Scores for Seniors

| | Partial Correlations with CLA found by Carini, Kuh & Klein (2006) | Partial Correlations found by Hosch (2010) |
|---|---|---|
| Minutes spent on CLA | -- | 0.33 ** |
| NSSE Benchmarks | | |
| Academic Challenge | 0.10 ** | -0.06 |
| Active and Collaborative Learning | 0.02 | -0.02 |
| Student-Faculty Interaction | 0.01 | -0.01 |
| Enriching Educational Experiences | 0.02 | 0.09 |
| Supportive Campus Environment | 0.13 *** | -0.10 |
| Self-Reported Gains in Learning Outcomes | | |
| Using computing and information technology | -- | 0.33 ** |
| Understanding yourself | -- | 0.29 ** |
| Analyzing quantitative problems | -- | 0.26 * |
| Thinking critically and analytically | -- | 0.25 * |
| Acquiring job or work-related knowledge and skills | -- | 0.24 * |
| Working effectively with others | -- | 0.12 |
| Learning effectively on your own | -- | 0.10 |
| Speaking clearly and effectively | -- | 0.08 |
| Writing clearly and effectively | -- | 0.05 |
| Acquiring a broad general education | 0.10 *** | 0.05 |

*p<0.05, **p<0.01, ***p<0.001; all tests were two-tailed. Partial correlations control for gender, enrollment status and SAT scores.

When controlling for SAT scores, gender, and enrollment status, correlations with CLA scores were observed with the number of minutes spent taking the test (R=0.33) as well as self-reported gains in learning outcomes in the areas of using computing and information technology (R=0.33), understanding oneself (R=0.29), analyzing quantitative problems (R=0.26), thinking critically and analytically (R=0.25), and acquiring job related skills (R=0.24). Perhaps most importantly, scores on the CLA for seniors in the present study did in fact correlate with their self-reported gains in thinking critically and analytically, which is a significant portion of the construct that CLA aims to measure. But it is also important to observe that this relationship is relatively weak (R=0.25 and $R^2$= 0.06). The NSSE benchmark relationships of Academic Challenge and Supportive Campus Environment with CLA scores that Carini, Kuh, and Klein observed were even weaker, and they were not observed in this study.

**Conclusions and Implications**

Linking results from CLA and NSSE over a period of three years at this public Carnegie Master's-Larger Programs institution demonstrated that in terms of engagement, students who took CLA were relatively representative of those who took the NSSE six months later. But the study also demonstrated that seniors – a much more difficult population to recruit – appeared to exhibit more engaged behaviors than students who did not take the CLA. In many ways, this finding appears a bit trite: students who are more involved in the life of the campus are more likely to come to campus to take a test in return for free graduation regalia than students who are less involved in the campus and may not even plan to attend the graduation ceremony. What is important about the finding, however, is that despite the appearance of representing the overall graduating class by race, gender, and field of study, the group of students who actually took the CLA may have different characteristics in terms of engagement, motivation, and drive that could influence their performance on the test. If motivation and time on test influence performance (Steedle, 2010; Hosch 2010) , then efforts to use CLA or other instruments for the purposes of accountability may not be able to rely upon institutional recruitment procedures to yield a sample of students who truly represent institutional performance.

Conversely, among seniors who took both CLA and NSSE, the correlations between test performance and the various NSSE engagement benchmarks was not observed in this study. Even in the previous multi-institution research of Carini, Kuh, and Klein, these correlations were weak at best. Some item-level correlations were observed, and in this respect, it is somewhat valuable to have confirmed some correspondence between institutional contribution to students' development in critical and analytical thinking and their CLA scores. Yet, while this connection was established, the relationship between the two variables only account for 6% of variation between the two of them. At this level of relationship, to what extent is it a valuable activity for an institutional research to administer three large surveys and six (sometime controversial) large test initiatives to generate a finding of this sort? There is some possibility that were the survey and the test more universal activities, then more valuable information might be gleaned, but current participation rates in these sorts of activities fall far short of universality.

Further consideration of potential flaws or limitations in each of the instruments is also likely warranted. Porter (2009) challenged the validity of NSSE and other higher education survey instruments, and called the field of higher education research to raise itself to a more rigorous level of validity. The weak correspondence between CLA and NSSE observed in this study and others may be indicative of this validity issue.

A more tantalizing consideration would be to examine with substantial rigor and method the extent to which CLA scores and NSSE results *should* correlate, especially because in this instance they were administered at different points in time. A common assumption behind almost all educational research is that students try their hardest when then take a test, yet this is patently not the case. A broad discussion of what test results (not just those from the CLA) are expected to mean. Why are such results superior to examination of a longer term work product—such as a senior thesis—that is more integrated into a curricular structure? As American education pursues better educational outcomes, the substitution of tests for research papers, theses, and long-term projects may place emphasis on an apparently quick and seemingly inexpensive measurement yet not develop the intended educational outcomes. And even worse, the testing regimen may detract from more effective and robust assessment practices that will more effectively advance the attainment of educational outcomes as well as provide actionable information about the extent to which they were achieved.

**References**


Arum, R., & Roksa, J. (2008). *Learning to reason and communicate in college: Initial report of findings from the CLA longitudinal study*. New York: Social Science Research Council.

Banta, T., & Pike, G. (2007). Revisiting the blind alley of value-added. *Assessment Update , 19* (1), 1-2.

Borden, M. H. & Young, J. W. (2008). Measurement Validity and Accountability for Student Learning. *New Directions for Institutional Research, Assessment Supplement 2007*.

Clarkberg, M., Robertson, D., & Einarson, M. (2008). Engagement and student surveys: Nonresponse and implications for reporting survey data. Paper presented at the meeting of the Association for Institutional Research, Seattle, WA. Retrieved November 11, 2010 from http://dpb.cornell.edu/documents/1000404.pdf.

Commission on the Future of Higher Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: U.S. Department of Education.

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

Garcia, P. (2007, November 14-16). *How to assess expected value-added: The CLA method*. Paper presented at the meeting of the California Association for Institutional Research, Monterey, CA.

Hosch, B. (2010, May 30 – June 2). *Time on test, student motivation, and performance on the Collegiate Learning Assessment: Implications for institutional accountability*. Paper presented at the meeting of the Association for Institutional Research, Chicago, IL.

Klein, S., Shavelson, R, Benjamin, R. (2007, February 8). Setting the record straight. *Inside Higher Ed.* Retrieved May 27, 2010 from http://www.insidehighered.com/views/2007/02/08/benjamin

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). Collegiate Learning Assessment: Facts and Fantasies. *Evaluation Review 31*, (5), 415-39.

Klein, S., Liu, O. L., & Sconing, J. (2009). Test Validity Study (TVS) Report. Voluntary System of Accountability. Washington, DC: Association of Public Land-Grant Universities.

Kuh, G. D. (2003). National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. Retrieved September 10, 2007, from National Survey of Student Engagement: http://nsse.iub.edu/pdf/conceptual_framework_2003.pdf.

Kuh, G. (2006). *Director's message in: Engaged learning: fostering success for all students*. Bloomington, IN: National Survey of Student Engagement.

McPherson, P., & Shulenburger, D. (2006). *Toward a Voluntary System of Accountability (VSA) for public universities and colleges*. Washington, DC: National Association of State Land Grants Universities and Colleges.

Miller, M. (2006). *Assessing college level learning*. San Jose, CA: National Center for Public Policy and Higher Education.

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

Steedle, J. (2009). Advancing institutional value-added score estimation. CLA White Paper. New York: Council for Aid to Education. Retrieved May 20, 2010, from http://www.collegiatelearningassessment.org/files/advancinginstlvalueadded.pdf.

Steedle, J. (2010, April 30 - May 4). *Incentives, motivation, and performance on a low-stakes test of college learning*. Paper presented at the meeting of the American Education Research Association, Denver, CO).

# HIGH STAKES FOR INSTITUTIONS, LOW STAKES FOR STUDENTS:

# STUDENT MOTIVATION AND STANDARDIZED ASSESSMENT

Thomas P. McGuinness

University of Michigan

Paper Presented at the North East Association for Institutional Research Annual Conference

in Saratoga Springs, New York

November 15, 2010

## Introduction

In recent decades, there has been a growing emphasis on accountability in higher education. The 1990s were marked by a movement toward performance based appropriations. States began to tie funding for public colleges and universities to sets of performance outcomes. Over the past few years, greater emphasis has been placed on accountability through measuring student outcomes. The federal government and accrediting bodies have recommended that institutions produce evidence of student learning. Former Secretary of Education Margaret Spellings' Commission on Future of Higher Education indicated in its report, "Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a "value-added" basis that takes into account students' academic baseline when assessing their results" (Spellings, 2006). Such recommendations are incredibly important to colleges and universities as there is concern that accreditation or even federal funding could be at stake. "Events in the wake of the Spellings Commission leave higher education in no position to simply wait until times change for the better" (Ewell, 2008). As a result, institutions, if they had not already, have implemented instruments to measure the value-added achievement of their students.

A major concern with the systematic value-added assessment taking place on campuses is that, while the stakes are incredibly high for institutions, the stakes are virtually non-existent for participating students. For example, a student can participate in the assessment without any preparation and without any incentive to perform well other than for the benefit of the institution. Additionally, students are often offered incentives to participate, suggesting that students would be unlikely to participate without compensation and that remuneration could be as much a motivator as an internal drive to perform well. Meanwhile, from an institutional perspective, a college or university needs students to take the test seriously so that it can illustrate that there are measurable gains in student learning that can be attributed to the education that the institution provides.

Compensating students for their participation is a factor only in a student's decision to participate. Once they are in the door, students can put forth as little or as much effort as they please. What is their motivation to perform on a low-stakes assessment? Performing well requires effort and can be time consuming. If there are no consequences, there is little incentive for students to put in anything greater than minimal effort. If the lack of stakes influence the level of effort put forth by students, the validity of the test is compromised. Yet, institutions rely on the data produced by these studies to guide their activities and potentially inform accrediting and government bodies about their students' levels of achievement. The aim of this study is to develop a greater understanding of how students are motivated to perform on tests that are low stakes for the student but high stakes for the institution.

## Literature Review

The challenges associated with lack of motivation to try hard on low stakes tests have been recognized by higher education researchers (Ewell, 1991; Palomba & Banta, 1999; Erwin & Wise, 2002). Erwin and Wise (2002) state, ''the challenge to motivate our students to give their best effort when there are few or no personal consequences is probably the most vexing assessment problem we face." Low student effort and motivation are a legitimate concern in determining whether low-stakes assessments are a valid measure of student achievement (Napoli and Raymond, 2004). Ewell (2006) revealed that the University of Texas system faced difficulties consistent with common criticisms of low stakes assessments. The experience in Texas using the CLA showed that the assessment produced data, though the data was highly suspect. "The UT testing initiative encountered familiar implementation difficulties in obtaining usable student CLA results, probably due to difficulties with student recruitment and motivation. For example, freshman scores are provided for only six of the nine UT campuses and senior scores for only seven because there were insufficient data for some campuses. Indeed, careful footnote readers of both reports will note the many caveats on the use of CLA results and strong warnings that not too much should be made of any given data point."

There are surprisingly few research studies on test taking motivation. The existing body of literature on the topic is fairly limited and is a product of the last two decades. At the K-12 level, Karmos and Karmos (1984) examined performance on the Stanford Achievement Test by students in grades 6 through 9 and found a significant correlation between student motivational attitudes and test scores. O'Neil, Sugrue, and Baker (1995/1996) tested the relationship between motivation and math achievement among 8$^{th}$ graders. Students were randomly assigned to one of

four groups: a financial incentive group (students received $1 for every correct answer), an ego instruction group (students were told the goal was to compare students), a task instruction group (students were told the goal of the test was to provide opportunity for personal accomplishment), and a control group. Students in the financial incentives group scored significantly higher on the exam and reported significantly higher levels of effort compared to the other three groups. There were no significant differences between the other groups.

O'Neil et al, in the same 1995/1996 study, found that there were no significant differences among treatment and control groups among 12th graders under the same research conditions. This raises the question of whether there is a meaningful interaction with age and whether the differences based on motivation persist to the undergraduate level. The research on motivation in postsecondary education, though, supports the hypothesis that stakes and motivation matter in testing. Wolf and Smith (1995) randomly assigned students to either a consequence group or a no consequence group on the initial day of testing. On the second testing day, students switched groups so that all students participated in both groups. The consequence group was told that their performance on the test would be counted as part of their course grade, while the no consequence group was told that the test would not count as part of their course grade. The consequence group reported significantly higher levels of effort and achieved significantly higher test scores than the no consequence group. Similarly, Napoli and Raymond (2004) examined the differences in performance on graded and ungraded exams among community college students and found that students taking the graded exam scored significantly better than those taking the ungraded exam. Sundre and Kitsantas (2004) found stakes to be a critical factor in examining the relationship between motivation and test performance. Among students who took a test with something at stake, there was no significant difference between the students who self-reported a high level of motivation and those with a low motivation level. However, when students took a test with no consequences, there was a significant difference in the scores based on self-reported motivation level. Their results showed that the standard deviation of test scores was much higher within the no consequences group, indicating a greater dispersion of scores when there is nothing at stake. Cole, Bergin, and Whittaker (2008) surveyed students who had completed the CollegeBASE standardized general education exam. The 26-question survey asked students to evaluate their experience based on interest, usefulness, and importance. The authors found that perceived usefulness and importance significantly predicted test-taking effort and performance. The study produced two useful findings. First, students who report trying hard on low stakes tests score higher than those who do not. Second, if students do not perceive importance or usefulness of an exam, their effort suffers and so does their test score.

## Conceptual Framework

The conceptual framework used in this study was chosen based on the hypothesis that students choose to perform on low stakes assessments due to their own altruistic characteristics. The work guiding the study is Richard Titmuss's *The Gift Relationship*. First published in 1970, *The Gift Relationship* made a case for systems of voluntary blood donation. Blood donation is sometimes described as a perfect example of altruism. The cross-national study examined institutional influences on variation in the blood supply. Titmuss compared the American and British systems, when it was legal to sell one's blood in the United States, and found that voluntary donation is both more socially just and economically efficient than the for-profit

exchange of blood. The book would eventually lead to a policy change in the United States in 1974 that would prohibit the commercial collection of blood (Healy, 2000). Based on this framework, this study investigates the correlation of motivation to perform and engagement in other altruistic activities, the impact of varying compensation on performance, and the relationship between the motivation to perform and interest in the well-being of the institution. This line of questioning determines whether students perceive their effort on their test to be an altruistic endeavor and whether changes in compensation would have any positive or negative impact on their motivation to perform.

## Methodology

In order to gain a better understanding of student motivation, this study employs an interpretive qualitative methodology in which students who have participated in a low-stakes standardized value-added assessment were asked to interpret their experiences. In fall 2008 and spring 2009, the Collegiate Learning Assessment (CLA), Measure of Academic Proficiency and Progress (MAPP), and Collegiate Assessment of Academic Proficiency (CAAP) tests were administered to approximately 1,100 students at thirteen colleges and universities in the United States as a part of the Voluntary System of Accountability's (VSA) Test Validity Study (TVS). The aim of the test validity was to determine the face validity and the construct validity of these three assessment instruments, which were recommended by a VSA task force. At each institution, 46 first-time, full-time freshmen and 46 seniors who had entered the institution as freshmen were recruited to take the tests. Students who completed the entirety of the study in three separate testing sessions were compensated with a $150 Amazon.com gift certificate (Shulenburger, 2009).

The data collected in this study consists of interviews with six sophomores at the University of Michigan (U-M) who participated in the TVS as first-year students. Due to restrictions with the U-M Institutional Review Board, I was unable to contact these students directly. Students who had participated in the TVS were contacted via email by a liaison in the Office of Budget and Planning. The message explained that I was a graduate student in the School of Education conducting research on assessment tests and that I was interested in hearing about their experiences. Interested students could then contact me to schedule an interview. Students were initially recruited without compensation, but, after failing to garner interest, participants were offered an incentive of $20 for the interview.

The interviews were conducted in a semi-structured format (see the Appendix for the interview protocol). They addressed several aspects of the students' experiences with the TVS. First, students were asked several descriptive questions about the process of participating in order to aid recall of their experiences, since they had participated in the TVS about a year prior to the interviews. The interviews then addressed students' motivations to participate in the TVS and to perform on the tests. Finally, students were asked to discuss their participation in charitable activities to test the hypothesis that a student's altruistic characteristics were a motivating factor in his or her decision to perform on the tests. In addition to asking about their charitable activities, informants were asked to what extent pride for U-M and service to the institution influenced their decision to perform or not on the TVS tests.

Each interview was transcribed after it had been completed. After all interviews had been completed and transcribed, the transcripts were coded first using an open coding approach to explore the descriptive themes that emerged in the interviews. These codes were then analyzed using an axial coding approach, which aims to relate descriptive themes to each other and put them into categories related to the topic area. Finally, using a selective coding approach, I organized the codes into major categories (Corbin & Strauss, 2009). Additionally, I wrote a brief memo after completing each interview.

## Limitations

The study had several limitations. First, the timing of the study relative to the TVS was problematic. Students were not interviewed until approximately a year after they participated in the TVS. At that point, they were asked to reflect on an experience that they had a year earlier. Over that period, it is likely that not only memories of details about the experience changed but also attitudes and interpretations. Optimally, they would have been interviewed immediately following their participation in the TVS. Second, I did not have the ability to triangulate the data with other sources. In particular, it would have been incredibly helpful to be able to factor the students' test scores into the analysis. Unfortunately, I was not allowed access to these data. Having test scores would have also been useful in determining to what extent I could trust the responses of informants. Students who indicated that they put forth effort on the exams may have been providing a more socially acceptable response. Additionally, the scores were not reported to students either, so even self-reported scores were not a possibility. Third, the small sample size from only one class at one institution is a threat to the external validity of the study. It is difficult to generalize these findings to the overall population. It is reasonable to assume variation not only across institutions but also within institutions, comparing the experiences of first-year students and seniors.

## Findings

The interviews revealed some common themes among informants about both their decisions to participate and their decision to perform on the tests. Within the broader category of deciding to participate, there were two themes that emerged within the category. The first was that *students use a calculated decision-making process to determine whether to participate*. The second was that, *due to the many competing research project demands on campus, undergraduate students must be offered an incentive in order to differentiate a project and attract their attention*. Within the broader category of deciding to perform, it was revealed that the hypothesis of the study was incorrect; participating students were not motivated to perform based on their altruism. Rather, the interviews revealed that *students are motivated to perform based on two factors: (1) respect for research and (2) personal pride.* Additionally, while students indicated that they put forth effort, they acknowledged that *they approach the tests differently from other tests because the stakes were low*.

### Decision to Participate

The factors that influenced students to participate in the study were discussed in detail in the interviews. While participating students demonstrated intellectual curiosity and expressed an interest in supporting research, each informant indicated that the monetary incentive was the

study component that initially caught his or her attention and was a critical part of the decision to participate. The interviews further revealed a deliberate decision making process, in which students calculated a value of their time versus the compensation offered by the study and considered the non-monetary value of their alternatives. Additionally, the interview subjects indicated that offering an incentive to undergraduates has become a necessity as a result of a saturated culture of evidence.

The interview subjects consistently demonstrated that they used a calculated process to determine whether the incentive offered in the TVS was sufficient to compensate them for their time. Allen, a biology major from Minnesota, was the most explicit in his description of his decision making process. When asked whether he still would have participated at varying levels of compensation ($100, $50, and no compensation), he talked through his thought process about whether he would decide to participate at each level. If offered $100, Allen responded, "Hmmm… each test was an hour and a half and there was three of them, which is about four and a half hours for $100? Yeah, I still would have done it for $100." Similarly, based on the incentive amount and the time commitment, regarding an incentive of $50, he replied, "Uhhh, four hours? I probably would have done it for $50. That would have been the end." For no compensation, though, "Uh, oof, no I would not have, not with this one." Allen went on to confirm that his decision making process was based on a deliberate mathematical calculation. He explained:

> "I guess I was figuring out how much I made each hour and, you know, there's a point when my time is worth more than the money."

Such a process was a common thread across interview subjects, though some students took the decision-making process a step further by verbalizing that they considered non-monetary alternatives. The opportunity cost associated with the utility derived from additional time devoted to schoolwork and even time spent socializing with friends was a consideration beyond the simple exchange of money for time. Brian, a philosophy major from Utah, explained in detail the factors influencing his decision to participate:

> "I kind of mentally just established a floor of, if it's below this amount then I won't participate because it doesn't actually increase my purchasing power by enough to justify the time commitment but if it's above this amount then there are some goods that I might want to purchase at varying levels for varying levels of compensation…If I weren't paid I wouldn't have participated, but beyond that I think I would've taken it somewhat less seriously. I think even a token amount of money would have sufficed. If I weren't offered anything at all, I would kind of approach it as though—obviously doing research costs money and so compensation can't always be offered—but for me it would just, especially given the large time commitment, it would register much lower on my list of priorities. I would think, well, I'm going to do this and there's some abstract benefit of going to do it but I'm not studying for my next exam or I'm not hanging out with a friend or whatnot and so even just having some sort of more tangible material benefit can then kind of allow me to say, 'Well, yes, I'm not studying for this but I

> am receiving something in return for the time.' Hence why I think a token amount even would have been sufficient."

Undergraduate students are incredibly busy, with competing academic, professional, social, and monetary interests, while time is a scarce resource. Interview subjects demonstrated that they are sensitive to balancing these interests and are careful in placing a value on their time.

Another important theme that emerged from the interview content related to the decision to participate was the fact that some sort of incentive is necessary to influence undergraduate students to participate in research. Particularly at a research university like U-M, students are bombarded with opportunities to support research projects. As they walk across the Diag, students are frequently approached by people conducting studies, while fliers about research opportunities are stuck to most lampposts and bulletin boards around campus. When students access their email, their inboxes are regularly flooded by requests to complete surveys or assist research efforts. In order to get their attention, students must be offered some sort of benefit in exchange for their time. Incentives with monetary values attached to them are the offers that are most obvious. When asked whether she would have participated had she received an email asking her to take part in the Test Validity Study without offering compensation, Rachel, a psychology major who is active in her sorority, responded:

> "Probably not, because I get hundreds of emails like that all the time saying—I get so many for Greek life, too, 'Greek life needs your help, like, Complete this survey,' or something like that and you get so many of them there's no way you're going to do all of them. I think that's the whole point of the compensation, to like stand out, because literally I get five or ten emails a day asking me to do something and I just don't have enough time in the day to do all of them so I'm going to do the ones that are most appealing. And a lot of them will say, 'University of Michigan Needs You' or 'Department of Psychology Needs You' or something like that."

While students may eventually derive intangible, non-monetary benefits, such as personal satisfaction and knowledge acquired, from their participation as subjects in research projects, it appears that it may be necessary to offer undergraduate students an incentive with an explicit monetary or material value. Otherwise, requests for support in research efforts are likely not to be differentiated and will find themselves in the trash as soon as they enter an undergraduate student's inbox.

**Decision to Perform**

In the interviews, it was clear that the informants knew exactly why they decided to participate in the TVS and that motivation was predominantly a result of the incentive offered. Students seemed less sure about why they actually put forth effort and perform on the tests; it was not a conscious decision-making process like it was when determining whether or not to participate. After thinking and talking through their decision-making process, informants frequently concluded that they were motivated by one or both of the following factors: (1) respect for research and (2) personal pride.

The majority of the students interviewed indicated that their motivation to perform on the tests was a result of their respect for research. The informants tended to recognize a moral imperative in being diligent in aiding research. They knew that if their performance was not an accurate reflection of their true abilities that it would negatively impact the quality of the research. One informant described well the idea that a few other students were not able to express as articulately:

> "I would judge someone negatively if they were to sign up for some study but then merely went through the motions rather than actually engaged in the study. Since, you know, setting aside factors of compensation which may motivate me or not motivate me to participate if I ultimately decided on participation then I have a certain moral responsibility to aid the researchers in collecting their data."

Another student echoed this sentiment and added that, while the exercise may have little meaning to participating students, the importance of the research should not be overlooked, "I'm sure it's important. I believe everything's important for the most part if people are going to spend their time on [research], there's got to be a reason." One student brought the idea to a more personal level, indicating that her motivation to put forth effort was out of respect for not only research but the researchers themselves. When asked what her primary motivation was, she replied, "I think just out of consideration for the researchers. I mean, I don't like to waste people's time, so that's probably the main reason why." Research is a major part of U-M and these students, even in their first year, had developed a respect for it. Informants determined that supporting research activities properly is the right thing to do.

The second factor that was a driving force behind the motivation of several of the students to perform was their personal pride. Vincent is a shy aspiring doctor from nearby Rochester. When asked why, when he could have put in minimal effort, he was motivated to try hard on the tests, he responded, "I don't know. I guess I always want to try my best even if it's not something that's going to affect my grade or anything." This is a sentiment that came up several times over the course of the interviews. U-M is a highly selective institution that attracts motivated students. They indicated that they put in the effort on the TVS tests because that is what they do in all of their endeavors. Said one informant:

> "I don't know. I think it's my morals maybe. Um, like I know people who, on tests, it'll be like, there's ten quizzes and the last one's dropped and if you got 100s on the nine of them, you could go in and not even answer the 10th, I would never be able to do that. I don't know why. I just wouldn't. Um, I mean, I don't always get 100% but I would never just, just not do anything or like not try at all."

While it is an easy choice to put in a minimal amount of effort on low stakes exams, high performing students are not accustomed to putting in the minimal amount of effort. They exhibit a personal pride by which they always try to do their best, even when there are not consequences.

There was another theme related to performance that emerged from the interview data. Each of the informants indicated that he or she put forth effort on the tests. However, they also acknowledged that they approach the tests differently from other tests because the stakes were

low. All six of the informants expressed that the tests not having any stakes for the student had an impact on their effort

- "I tried to get into the mindset that the test was important for my career, my future, whatnot. Although of course I mean obviously I knew in the back of my head that it wasn't. And so I'm sure that had some impact on level of effort I gave, but I generally think that I performed roughly as I would have if it were real test."
- "There was a writing thing where we had to do like three different writing essays and, like, the person next to me literally took like 10-15 minutes to do all of them. So, I mean, I think it was bad for them to say that, though clearly, if we're getting paid, there's not much implication on our grade or anything, but some people took it much more seriously than others. Yeah, my roommate also did it and he told me that he really didn't try."
- "There was a time limit but I knew that if I don't know I just felt comfortable really just trying to think through it, um, because I did still want to try to get the answer right but I wasn't totally too worried about it."
- "I mean, for the hard questions I probably didn't make as thoughtful an answer for my final exam in a class but I didn't just completely fudge it. So, there wasn't as much pressure to do well because I know it wasn't really a reflection on me, personally, besides for their research purposes, so it wasn't the pressure but I tried to give a semi-thoughtful answer, at least."
- "I definitely tried my best but maybe in some aspects if I knew it was being like, going to affect me, I would have tried harder… maybe I would have tried harder or, like, focus a little more maybe if there was like a question where you had to go back to the reading, I was just like, "I think I remember," and I would circle it, instead of, if I really need this grade, maybe I would have gone back and been like, I really need to check."
- "I definitely would have tried a lot harder. And I probably would have prepared somehow, maybe would have looked over some general things like maybe basic chemistry or basic math, just to make sure that I knew it."

Even these motivated students who reported that they put in effort on the tests acknowledged that their effort level was lower than it would have been on a test with stakes. This supports criticisms the validity of these tests is threatened because they are low stakes.

## Discussion

With the growing demand for accountability and the need to produce evidence of student learning, interest in standardized value-added testing is on the rise. No matter how much research is done on the reliability and validity of the testing instruments, if students do not put forth effort because nothing is at stake then the measures are flawed. We need a better understanding about student approaches to this type of testing, whether students are motivated to perform, and, if they are, what motivates them to perform. I believe that this study can be a useful addition to the

research on undergraduate value-added assessment. However, the study has clear limitations that need to be addressed before making any generalizations about the findings.

It comes as little surprise that students elected to participate in the TVS in large part due to the incentive offered. This generous incentive was offered to a population eager to be compensated. What was a more interesting finding was that, due to the many competing research project demands on campus, undergraduate students must be offered an incentive in order to differentiate a project and attract their attention. I suspect that this is a phenomenon more typical at a large research university like U-M, where there are a multitude of studies going on at any time and undergraduates are needed as participants. At a liberal arts college, where the scope of research is narrower, this may not be the case.

Regarding the students' decisions to put forth effort and perform on the tests, it was surprising that none of the informants made the connection between participating in a study attempting to measure institutional quality with service to the institution. Each of the informants described a commitment to charitable activities, yet none of them approached the test thinking that it could have any positive of negative consequences for U-M. It was refreshing, though, that they recognized that a lack of performance could compromise the integrity of the research. While it did not introduce anything particularly original, students' admissions that they approached the test differently because it was low stakes supported previous findings about motivation on low stakes tests. Each of the students indicated that he or she worked hard on the tests, though not at the level they would for a high stakes assessment. This is quite problematic when institutions have much riding on these tests.

There are certainly opportunities for future research. Most importantly, having test results would yield a much richer analysis. All of the informants indicated that they worked hard on the tests, yet several reported that others who took the tests put in a level of effort that seemed to be insufficient. This presumes that either the students with which I met provided me with socially acceptable responses or my sample of students did not include those who put in a low level of effort. Being able to triangulate the qualitative data with the quantitative results would help in determining to what extent students actually did make an effort to perform. Additionally, having access to such data might allow for an empirical study that could test the factors that influence a student's performance.

In addition to increasing the depth of data, it would also be useful to increase the breadth of data. Conducting additional interviews and expanding the study to other institutions is necessary in order to generate compelling conclusions. The current study, due to its small sample size, can generalize the experiences of first-year students in U-M's College of Literature, Science, and the Arts who participated in the TVS in fall 2008. We cannot generalize to seniors who take these tests, students at other institutions or colleges within U-M, or even students who participate in value-added testing in other administration cycles. There is likely to be variability across quite a few variables, including but not limited to institution type, academic class, academic discipline, and financial background. In order to truly understand student attitudes towards low stakes testing, these differences need to be examined.

## References


Cole, J. S., Bergin D. A., Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*, 609–624.

Corbin, J. & Strauss, A. (2009). *The Basics of Qualitative Research.* Thousand Oaks, CA: Sage.

Erwin, T. D., & Wise, S. L. (2002). A scholar-practitioner model for assessment. In T. W. Banta (Ed.), *Building a scholarship of assessment* (pp. 67–81). San Francisco: Jossey-Bass.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Education, 17*, 75–125.

Ewell, P. T. (2006). From the States: Smoke Signals from the Hills of Texas. *Assessment Update, 18(3), 12-15.*

Ewell, P. T. (2008). Assessment and Accountability in America Today: Background and Context. *New Directions for Institutional Research, Assessment Supplement 2007*, 7–17.

Healy, K. (2000). Embedded Altruism: Blood Collection Regimes and the European Union's Donor Population. *American Journal of Sociology*, *105*(6), 1633–1657.

Karmos, A. H., & Karmos, J. S. (1984). Attitudes toward standardized achievement tests and their relation to achievement test performance. *Measurement and Evaluation in Counseling and Development, 17,* 56–66.

Napoli, A. R., & Raymond, L. A. (2004). How reliable are out assessment data? A comparison of the reliability of data produced in graded and ungraded conditions. *Research in Higher Education, 45,* 921–929.

O'Neil, H. F., Sugrue, B., & Baker, E. L. (1995/1996). Effects of motivational interventions on the National Assessment of Educational Progress Mathematics performance. *Educational Assessment, 3,* 135–157.

Palomba, C. A., & Banta, T. W. (1999) Assessment essentials: Planning implementing and improving assessment in higher education. San Francisco: Jossey-Bass.

Secretary's Commission on the Future of Higher Education. 2006. A Test of Leadership: Charting the Future of U.S. Higher Education. Washington, DC: U.S. Department of Education.

Shulenberger, D., et al (2009, November) *Interpretation of Findings of the Test Validity Study Conducted for the Voluntary System of Accountability*. Retrieved from the Voluntary System of Accountability website: www.voluntarysystem.org/docs/reports/VSAabstract_TVS.pdf

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29,* 6–26.

Titmuss, R. M. (1997). *The Gift Relationship: From Human Blood to Social Policy, Expanded and revised edition*. Ann Oakley and John Ashton (eds.). New York: The New Press.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8,* 227–242.

# Appendix

Background Information

1. Introduce myself and explain the study.
2. What was it that influenced you to choose the University of Michigan?

Becoming Involved in the Assessment Experience

3. Last year you took the CLA, MAPP, and CAPP tests. Could you tell me how you became involved with this project?
4. Who was involved in the recruitment and administration of the tests?
    a. What background information did they give you about the tests?
    b. Did they mention anything about what was at stake in taking the tests? If so, what did they tell you?
    c. What did they tell you about how you should approach the tests?

Motivation to Participate

5. How did you decide to participate?
    a. If money was a factor:
        i. If you were only paid $50, would you still have participated?
        ii. If you weren't paid, would you still have participated?
    b. What were your expectations of the experience?

Motivation to Perform Well

6. In what ways did you prepare for the test?
    a. In retrospect, would you have prepared differently? If so, how?
7. Do you think the compensation offered had any impact on your performance on the test?
    a. If you weren't paid, would that have impacted your effort?
8. Would you have approached the test differently had it been tied to something high stakes, such as a grade or a graduation requirement?

Other Forms of Altruism/Motivation

9. Can you tell me about any charitable activities in which you participate?
    a. Community service
    b. Blood donation
    c. Philanthropy
10. Would you say that you are engaged in the school spirit of U of M?
    a. In what ways?
11. I ask these questions because I'd like to know whether pride for the University of Michigan and service to the institution had any impact on your experience. Would you say that this is the case?

Additional Thoughts

12. Are there any other aspects of the experience that we haven't already covered that you would be willing to share?

**NEAIR 37TH ANNUAL CONFERENCE, NOV. 2010**

# HOW THE NEW RACE/ETHNICITY QUESTIONS CHANGED DEMOGRAPHICS OF LSAT TEST TAKERS

PHIL HANDWERK
INSTITUTIONAL RESEARCHER

LISA STILWELL
SENIOR RESEARCH ASSOCIATE

LAW SCHOOL ADMISSION COUNCIL

# **ETHNICITY** CATEGORIES

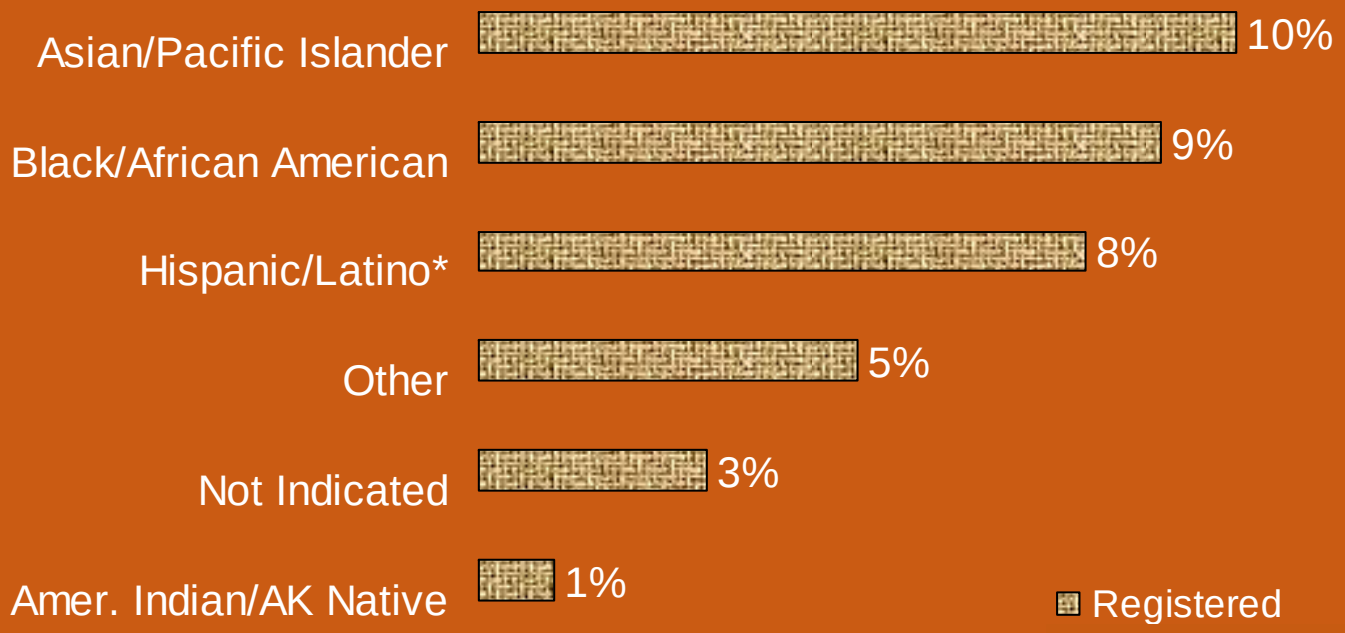
## PREVIOUS OPTIONS ON LSAT ANSWER SHEET

- American Indian/Alaska Native
- Asian/Pacific Islander
- Black/African American
- Canadian Aboriginal
- Caucasian/White
- Hispanic/Latino
- Chicano/Mexican American
- Puerto Rican
- Other
- Not Indicated

## NEW RACE/ETHNICITY CATEGORIES

- American Indian/Alaska Native
- Asian
- Native Hawaiian/Other Pacific Islander
- Black/African American
- Canadian Aboriginal
- Caucasian/White
- Hispanic/Latino
- Puerto Rican
- Aboriginal/Torres Strait Islander Australians
- Not Indicated

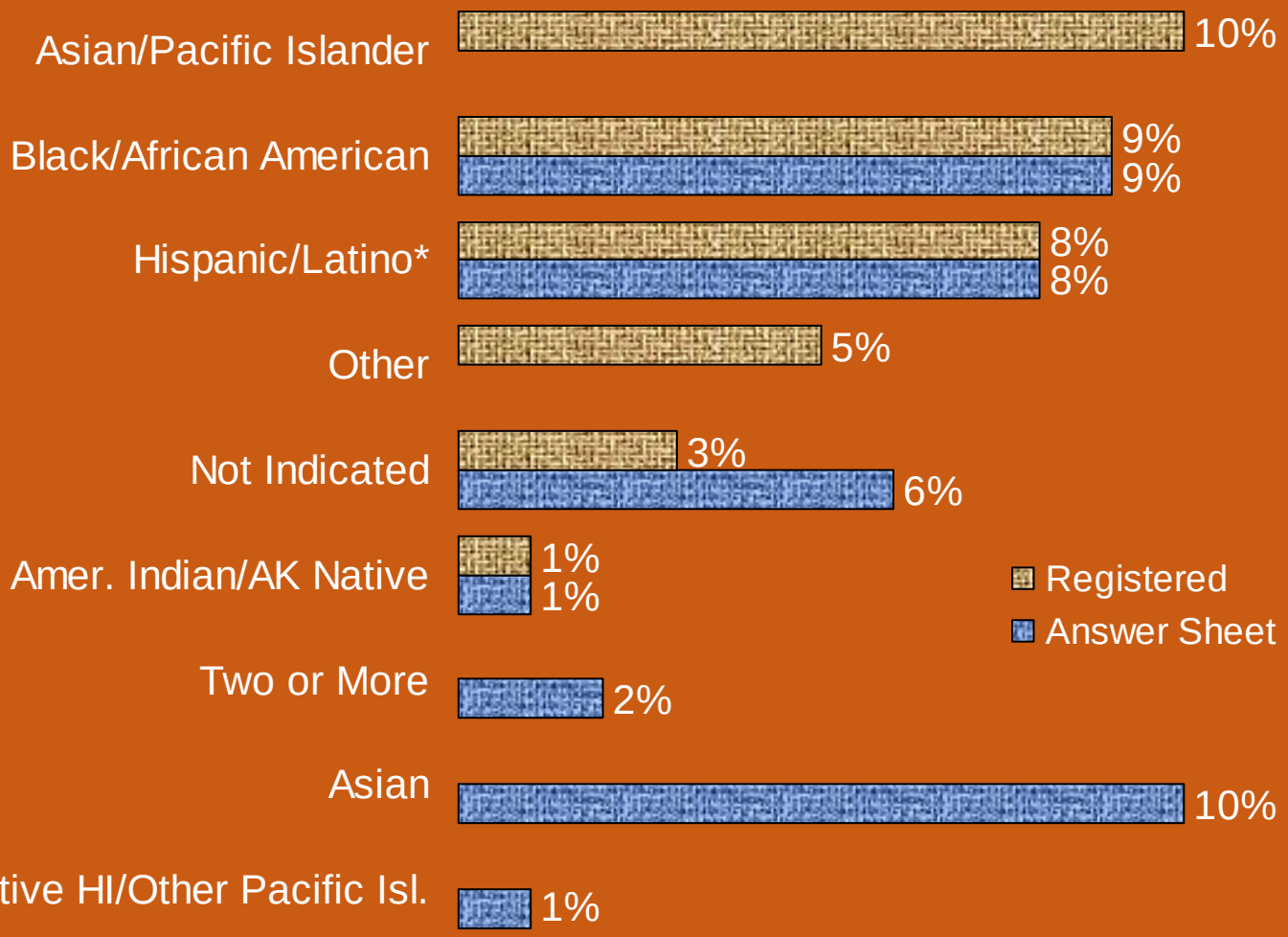
# **JUNE** 2009 LSAT TEST TAKERS BY RACE/ETHNICITY USING DEPT ED REPORTING




* Includes Chicano/Mex Am & Puerto Rican.

Caucasian/White: Registered – 63%

# **JUNE** 2009 LSAT TEST TAKERS BY RACE/ETHNICITY USING DEPT ED REPORTING

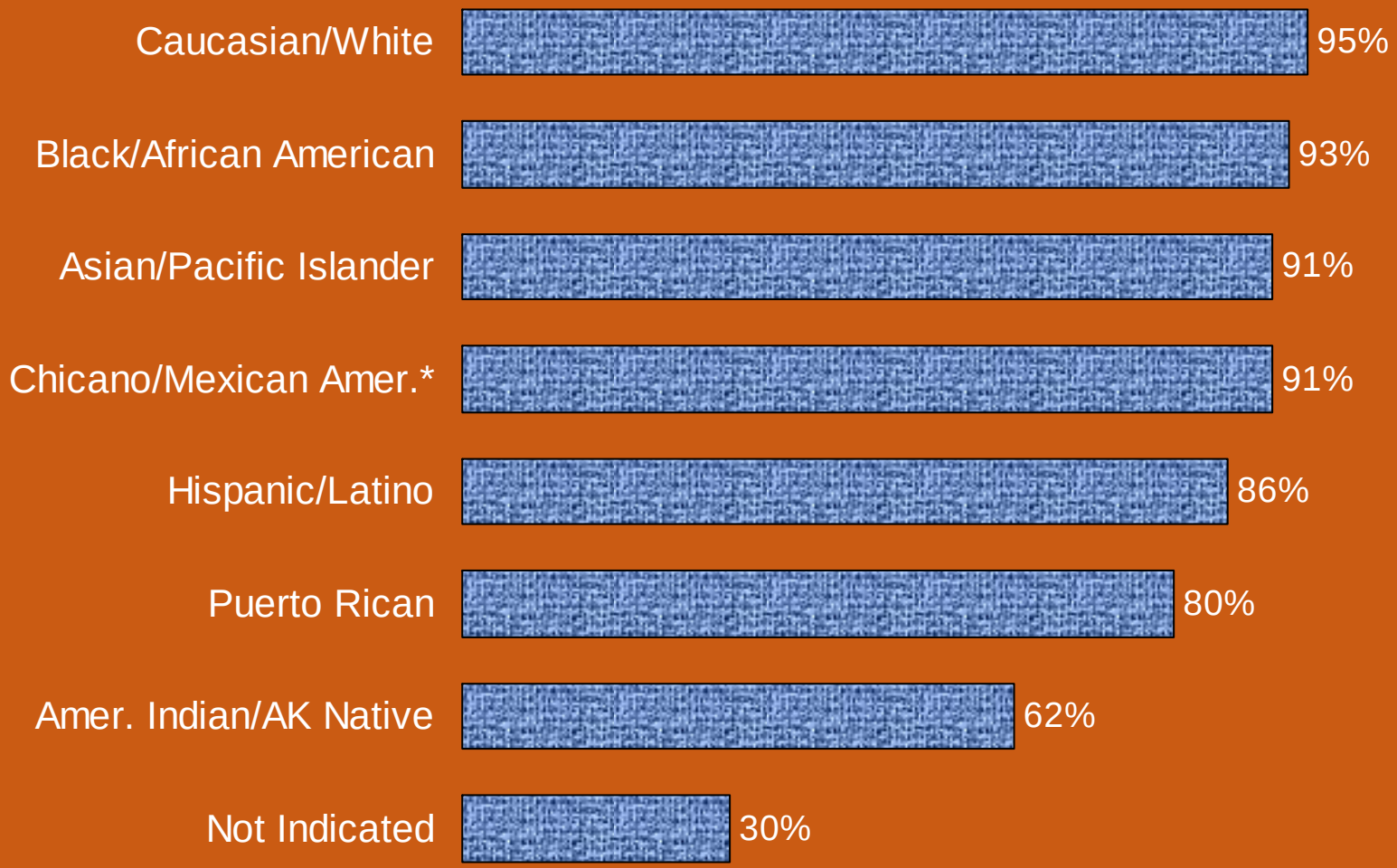



* Includes Chicano/Mex Am & Puerto Rican.

Caucasian/White: Registered – 63%; Answer Sheet – 64%

LSAC.

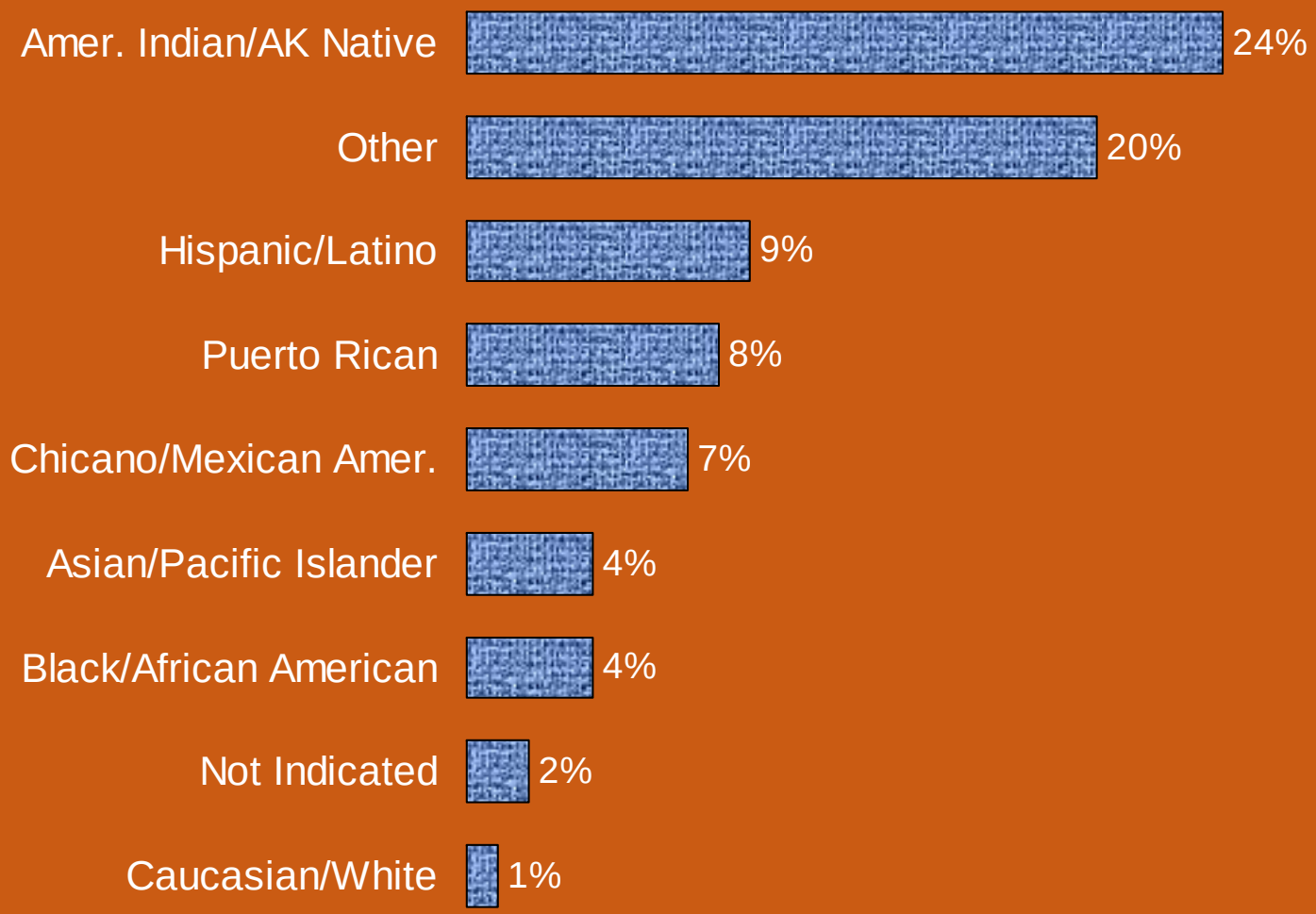
# **PERCENT** JUNE 2009 LSAT TEST TAKERS INDICATING SAME RACE/ETHNICITY ON ANSWER SHEET AS AT REGISTRATION



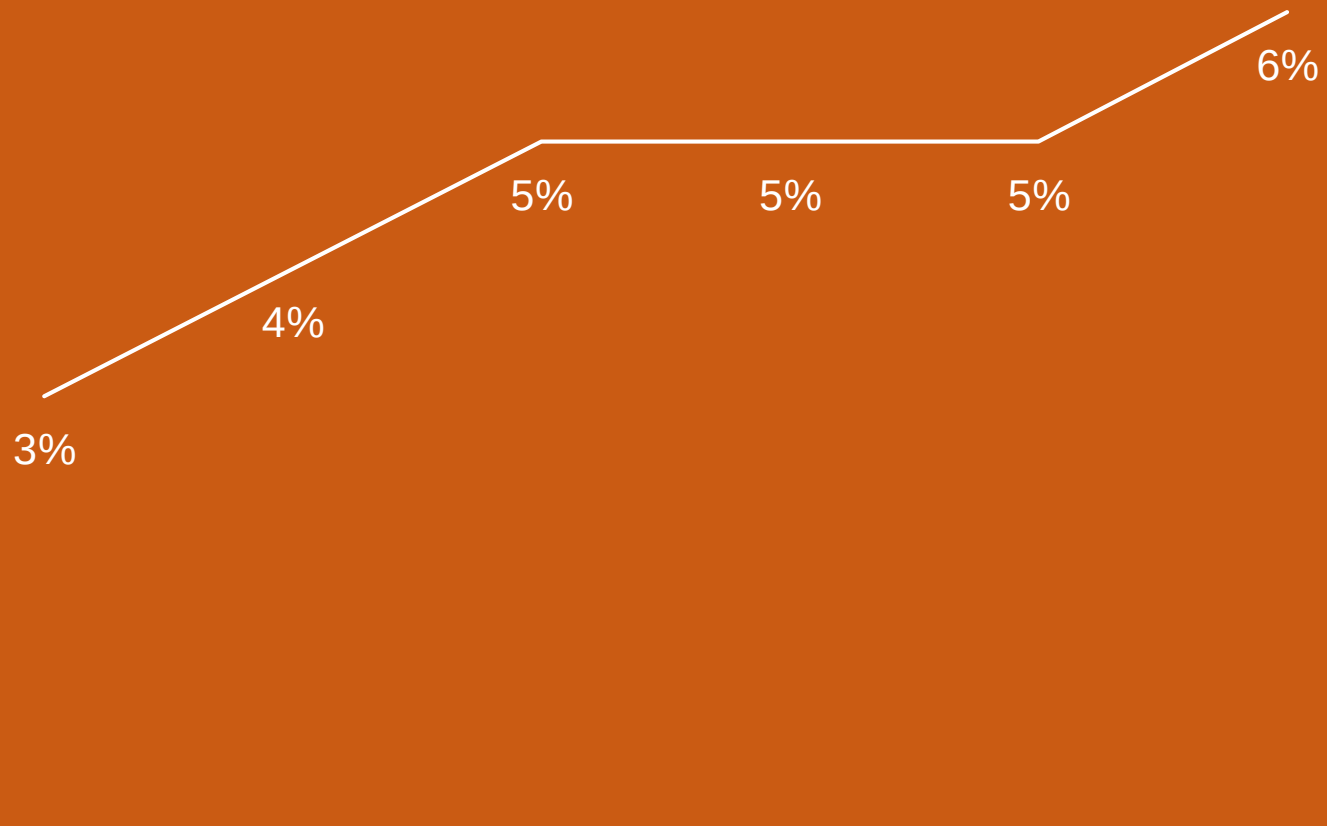

* Indicated Hispanic/Latino on answer sheet

LSAC.

# **PERCENT** JUNE 2009 LSAT TEST TAKERS WITH MULTIPLE RACE/ETHNICITY ON ANSWER SHEET BY REGISTRATION RACE/ETHNICITY



# **PERCENT** TEST TAKERS* INDICATING MULTIPLE RACE/ETHNICITY




* Based on main form only

# QUESTIONS?

PHANDWERK@LSAC.ORG

LSTILWELL@LSAC.ORG

# ISSUES IN WEB SURVEYS OF STUDENT POPULATIONS:

# RESPONSE RATES AND POST-STRATIFICATION WEIGHTING

Joel David Bloom, Ph.D.

Director of Assessment & Survey Research
Office of Institutional Research, Planning & Effectiveness

Research Assistant Professor
Department of Political Science

The University at Albany, SUNY
1400 Washington Ave.
Albany, NY 12222

jbloom@albany.edu
(518) 437-4791

Abstract. This research is in two parts, regarding (1) response rates and (2) weighting. For response rates, the objectives were to test whether changes to a number of experimental conditions would have an important impact on response rates in student surveys. I present results from 6 response rate experiments, finding that using personalized solicitations improves response rates, but changes in subject line and use of e-mail pre-notification do not. Regarding weighting, the objectives were to determine first the feasibility of applying post-stratification weights to student survey data, and then test whether the use of those weights would have a material impact on the results. An additional objective to the research on weighting is to determine whether a shift from use of paper surveys administered to a sample of classrooms to a web-based survey in which all students are invited to participate, would make a substantive difference to response patterns. After using post-stratification weighting to correct for differences between the student population and survey respondents using three different surveys, I find no impact but maintain that weighting still constitutes best practices in certain types of surveys and should at least be checked for surveys of high importance.

Note: Portions of this paper have been presented previously at the 2008 and 2010 Annual Meetings of the American Association for Public Opinion Research.

## I. Experimental Tests on Response Rates in Student Web Surveys:

## What Works and What Doesn't

Introduction: Why Response Rates Matter in Student Surveys

Over the last ten years, a plethora of response rate studies too numerous to cite individually has shown that, once some fairly minimal level of response rate has been achieved, response rates have much less of an impact on survey data reliability than do other factors, notably non-response bias and sample representativeness. In surveys of students conducted by universities, we have the advantage of having high quality, regularly updated comprehensive lists of our students with 100% population coverage and detailed academic and demographic variables that enable us to compare our survey samples with the overall student population with a level of precision that would be the envy of survey researchers dealing with more general populations. Thus, we can always check the sample against the population very precisely and easily weight the data if necessary (see, e.g., Bloom, 2008) to correct for sample non-representativeness and the potential non-response bias that can go with that.

Given all that, why should we care about response rates? We care about response rates due to a combination of two factors: (1) higher response rates lead to larger sample sizes, which in turn (a) lead to more precise estimates on individual questions and (b) enable us to break the sample down into more subsamples of interest, based on various demographic and academic groupings; (2) due to the large number of student surveys requested by various units of the university, obtaining higher response rates enables us to: (a) divide the population into samples, each of which will be invited to receive a different survey; higher response rates mean we can divide students into smaller samples and conduct more of them; and (b) send fewer contact e-mails, which in turn enables us to start another survey more quickly than if we had to send an additional follow-up e-mail and hold the survey open another week.

As alluded to before, university students are a population virtually ideally suited for web surveys. We have lists with 100% coverage, and we have e-mail addresses for almost all of them (while this of course does not guarantee that everyone uses or regularly checks their e-mail account, this problem is associated with mail and list samples of any populations). In addition, we have tremendous quantities of other administrative data that we can either pre-seed the survey data set with upon login or merge with the survey data after the fact, if the survey is not anonymous.

Unfortunately, due to financial constraints, we do not have the resources to engage in survey best practice that have been shown to workin numerous split-sample experiments.These include sending mailed pre-notificationsor mailed or telephone reminders for the web survey, and sending small up-front incentives to everyone, the latter of which has been shown repeatedly to be much more effective than offering a

lottery-type incentive to those who complete the survey (Couper, 2008; Dillman, Smyth and Christian, 2009). But since any of these options would require several thousand dollars in cost, and some would also require many hours of staff time, they are simply not an option here, or at many other universities.

As a result, our only options for improving response rates at the University at Albany are: (1) sending e-mails and (2) sending more e-mails. Thus, it is of great importance for us to determine how to make the best use of these e-mails to improve our response rates; because we must go back to the same population repeatedly, it is also of utmost importance that we do this without causing too much survey fatigue, risking "poisoning the well" for future surveys.

The use of web-based surveys utilizing e-mail invitations is new enough that it is only beginning to develop anexperiment-based literature on effective means of improving response rates. The most thorough review of this literature is to be found in *Designing Effective Web Surveys*, by Mick Couper (2008); the 3rd edition of *Internet, Mail, and Mixed Mode Surveys*, by Dillman, Smyth and Christian (2009) also provides an excellent summary of the current state of the field.

## The Research Questions

Among the most important issues related to improving response rates are (1) personalization of the e-mail invitations; (2) the content of the subject line; (3) the source of the e-mail as it appears on the "from" line; and (4) use of pre-notifications. In this paper, I present the results of a total of 6 experiments conducted on three web-based surveys conducted in 2007 and 2009 at the University at Albany, SUNY, a medium-sized research university in the Northeastern United States.

1. Personalization. The efficacy of personalizing invitation letters for mail surveys has long been recognized, and has been supported by decades of experimental research. As Dillman, Smyth and Christian explain,

> Social and behavioral scientists have long known that in emergency situations, the more bystanders there are, the less likely anyone is to step forward and help out…Although less dramatic, the goal of personalizing survey contacts is quite similar: to draw the respondent out of the group….Moreover, personalization can be used to establish the authenticity of the survey sponsor and the survey itself and to gain the trust of respondents, both of which should improve the likelihood of response. (2009, p. 237)

Dillman and his co-authors find that the same reasoning that had long been established for personalizing invitation letters in mail surveys applies equally well to web-based surveys:

> Personalizing all contacts in web surveys is important for the same reason as in mail surveys – it establishes a connection between the surveyor and the respondent that is necessary to invoke social exchange, and it draws the respondent out of the group. (2009, p. 273)

They list a number of studies showing that this is actually the case (Heerwegh, 2005; Joinson&Reips, 2007). Couper (2008) lists those two as well as several others (Porter and Whitcomb, 2003; Pearson and Levine, 2003; Joinson, Woodley, &Reips, 2007), each finding that the group receiving the personalized invitation had higher response rates than the group receiving the generic solicitation. In three of the experiments discussed in this paper, I tested whether a personalized invitation in two cases or pre-notification e-mail, would increase response rates among our population of university undergraduates.

2. Subject Line Content. Far less research has been done on the most effective use of subject lines. Dillman, et al., suggest that the subject line mention that the e-mail is about a survey, and that it include a request for assistance:

> The subject line should…be professional and informative. It should immediately tell the respondent that the e-mail is about a survey, who the sponsor is, and what the topic is….Consistent with the social exchange perspective, some research has found that stating the subject as a request for help rather than an offer to let students share their opinions results in increased response. (Trouteaud, 2004; cited in Dillman, op. cit. p. 286)

Couper sites two studies in which manipulations in subject line content regarding the purpose of the e-mail (a survey) and whether it was phrased as a request or an offer had little or no impact on response rates among university populations. (Porter and Whitcomb, 2005; Damschroder, unpublished). As Couper puts it,

> My guess is that the decision to open an e-mail message, especially from a known or recognized sender, is not a deeply processed one. Beyond some minimal threshold to verify that the sender is a known entity, and thus the e-mail is not spam, the subject line may receive relatively little attention. (Couper, p. 315)

In the experiments discussed below, we tested the use of the word “survey” against the request for “input” to help determine which of these factors, if either, would have the stronger impact on responses.

3. E-mail Sender. Another factor that, along with subject line content, has not been the subject of a great deal of research as yet, is the format and identity of the e-mail sender. As Dillman and his coauthors point out:

> Once an e-mail gets past spam filters and delivered into an inbox, the recipient generally has only two sources of information to use in determining whether to open the message; the text that appears in the “From” field and the subject line. As a result, these two pieces of information need to convince the respondent that this is an important message from a reputable sender. Thus, it is important to send the e-mail requests from a professional-appearing e-mail sender and address. (Dillman, op. cit. p. 285)

Coupermakes much the same point, adding that the survey researcher needs to take full advantage of the fact that these elements are often visible even without opening the e-mail:

> Given that the three header elements (sender, recipient, and subject) are often visible without opening the e-mail message, they should convey the importance of and legitimacy of the request….Enough information needs to be conveyed in the header to reassure the recipient, and encourage the opening and reading of the e-mail message. If that is done, more information can be conveyed in the body of the message. (Couper, pp. 315-316)

Joinson and Reips (2007) and Joinson, Woodley, and Reips (2007) found in their panel studies that e-mails sent by high-status senders received higher response rates than those of lower status, and that personalization was most effective if the sender is of high status. However, in most surveys conducted by my office, the sender is a high-status administrator such as a Vice Provost or Vice President; for our purposes, the bigger question was whether the e-mail really had to come from that person's own e-mail account, or whether it was sufficient to send it from a more generic account under that person's name. The results of a test of this question for a survey pre-notification e-mail are detailed below.

4. Pre-notifications. Pre-notifications have been shown to be important in improving response rates in mixed-mode surveys, especially when the pre-notification is sent in a different mode than the survey invitation itself. Examples would include a mailed pre-notification for a web-based survey, or vice versa. Crawford et al. (2004), Kaplowitz et al. (2005) and Dillman et al. (2009) all show experimental evidence that a mailed pre-notification can significantly improve the response rate of a web survey. However, whether an e-mailed pre-notification would improve response rates in a web-based survey is another question. As Couper notes dryly, "An e-mail prenotice…is likely to be less effective than a contact using another mode." (p. 306) In the final set of experiments presented below, we examine precisely that question.

## Experiments 1 and 2: The 2007 UAlbany Student Experience Survey (SES)

The Student Experience Survey (SES) is a comprehensive survey administered to undergraduates at the University at Albany every few years. The SES was specifically designed to be UAlbany's major quantitative tool for utilizing and further testing the "Albany Outcomes Assessment Model," first developed in the late 1970s, which seeks to demonstrate UAlbany's impact on students' intellectual, personal and social growth.The "Albany Model" includes four major components: (1) personal traits; (2) college experiences; (3) educational outcomes; and (4) alumni outcomes. SES questions cover a wide variety of issues related to UAlbany undergraduate students with regard to all four of these areas. Non-seniors were asked a total of 108 questions; graduating seniors were asked up to an additional 26.

In order to garner a sufficient response rate, we took a number of steps above and beyond what we do with less high-priority surveys. These steps included placing posters around campus, and requesting e-mails from academic advisors and department and

program chairs, as well as offering a chance to win one of five cash prizes of $50.00.[1] The first invitation e-mail was sent via the undergraduate student listserv (which appears as "Academic Affairs-Notices" on the "from" line) on Monday, March 19th, 2007 with the subject line: "UAlbany Student Experience -- your input needed" and the salutation "Dear UAlbany Student"and the signature of the Vice Provost for Undergraduate Education. The first reminder e-mail was also sent via the listserv on Tuesday, March 27th, with the subject line "An Important Message from SA President [name]," the salutation "Dear fellow UAlbany students" and the signature of UAlbany Student Association President.

Despite all this, when students returned from Spring Break, we still had onlya 14.5% response rate (1,660 responses out of a population of 11,424 matriculated undergraduates). At this time we decided to keep the survey open and send another reminder e-mail; while we were at it we decided to embed two split-sample experiments into this third e-mail to test two hypotheses: (1) that, consistent with the literature on mail surveys, personalizing the salutation would lead to an increased response rate compared to having a generic salutation as we had always done previously; (2) that including the word "survey" in the subject line might scare some people off and lead to a reduced response rate compared to subject lines that mention "input."

So on Thursday, April 12th, a final e-mail was sent out under the author's name and signature. The 9,735 students who had not yet completed the survey were divided randomly into four nearly equal groups so that one half of the studentswere sent e-mails using Microsoft Outlook's "blind copy" (bcc) function with the solicitation "Dear UAlbany Student" and the other half were sent a personalized "Dear [first name]" using Outlook's "mail-merge" function. The other experiment had to do with the subject line – Half of each previously-mentioned group received each of two slightly different subject lines:"Final Reminder: UAlbany Student Experience Survey" or "Final Reminder: UAlbany Needs Your Input."

As shown in Table 1a and 1b, below, personalization does help. From the two groups with which I used the mail-merge and a personal salutation, we received a total of 211 new responses. From the two groups with which I used the generalized "blind cc" method, we received 152 new responses. Thus, the personalization was associated with a 39% increase in the number of raw responses. Table 1b shows results of a difference-of-mean test in which the mean for each sub-sample is the response rate; as expected, the difference was statistically significant at a high level, with a t-ratio of 3.155 (p=.002). This is consistent with the literature discussed above showing increased response rate with personalized salutations.

[1]As discussed in the introduction, we are aware of the higher effectiveness of smaller up-front cash gifts as incentives, but lack the budget for them.

Table 1a. Response Rates with all Four Split-Sample Categories, SES 2007.

| | Personal/"Survey" | Personal/"Input" | BCC/"Survey" | BCC/"Input" | Total |
|---|---|---|---|---|---|
| Respondent Count | 101 | 110 | 76 | 76 | 363 |
| Non-Respondent Count | 2333 | 2324 | 2357 | 2358 | 9372 |
| Total Count | 2434 | 2434 | 2433 | 2434 | 9735 |
| Response Rate | 4.15% | 4.52% | 3.12% | 3.12% | 3.73 % |

Table 1b.Hypothesis Test of E-mail Personalization, SES 2007.

| | Personalized | Non-Personalized | Difference |
|---|---|---|---|
| Respondent Count | 211 | 152 | 59 |
| Non-Respondent Count | 4657 | 4715 | -58 |
| Total Count | 4868 | 4867 | -1 |
| Response Rate (Mean) | 4.33% | 3.12% | 1.21% |
| Standard Deviation | 20.37 | 17.40 | |

t = 3.155; df = 9733; sig (2-tailed) = 0.002

Table 1c.Hypothesis Test of Use of "Survey" in Subject Line, SES 2007.

| | "Input" | "Survey" | Difference |
|---|---|---|---|
| Respondent Count | 186 | 177 | 9 |
| Non-Respondent Count | 4682 | 4690 | -8 |
| Total Count | 4868 | 4867 | -1 |
| Response Rate (Mean) | 3.82% | 3.64% | 0.18% |
| Standard Deviation | 19.17 | 18.72 | |

t = 0.479; df = 9733; sig (2-tailed) = 0.632

In the other test, however, the difference in the subject line didn't matter. Overall, 186 students sent an e-mail with the word "input" in the subject line completed the survey, compared to 177 of those with the word "survey," a much smaller difference of only 5% increase in the number of raw responses. Not surprisingly, this difference, while in the expected direction, was not statistically significant, with a t-ratio of 0.479 (p=.632). Of course this does not mean that no differences in the subject line would matter, just that the two I tried had statistically indistinguishable results. However, these results are broadly in line with Couper's observation above that one would not necessarily expect the subject line to have a great impact when the e-mail is already from a fairly trusted and well-known source.

Experiment 3: The 2007 UAlbany Cable Survey

Later that same term, our office was asked to conduct a survey of students living on-campus regarding their opinions of and experiences with the University's in-house cable television channel. Having just received the results from the Student Experience Survey (SES) described above, we decided to do an additional split-sample experiment on survey personalization, in order to (hopefully) provide additional confirmation for the SES results. This was a much shorter survey, with only 14 questions, including one open-ended comments question. It was also a lower-priority "quick and dirty" survey, in contrast to the higher-priority, longer, more comprehensive SES, which had been in development and use (in current and earlier forms) literally for decades. In addition, because it came so late in the semester, the Cable Survey would only have a single invitation e-mail with no follow-up reminders, and the experiment would take place on this single (and thus first) invitation compared to the third e-mail on the SES. Because the two surveys were so different in so many ways, it would provide especially strong confirmation of the hypothesis if we were to find here as well that students addressed by name were more likely to take the survey than those addressed generically.

Both versions of the e-mail invitation were sent out on Friday morning, April 27th, 2007 with the subject line: "Your Input Needed on UAlbany Cable TV!" Because the e-mails were sent out from the author's e-mail account (as discussed below, we subsequently created a "UAlbany Survey" account for this purpose) we also included a line at the top stating: "The following is a special message from UAlbany Vice President [name]" under whose signature the e-mail was also sent. As with the SES, one group was sent the message by pasting their e-mails into the "bcc" box; this group was addressed as "Dear UAlbany Student." The other group was sent the same e-mail via mail-merge addressed to "Dear [first name]."

The results of this experiment are shown in Table 2, below. When the survey was closed on Monday, May 7th, 326 of the students who were addressed personally responded (for a 9.4% response rate), compared to only 265 of those addressed generically (for a 7.7% response rate).This translates to a 23% increase in the raw numbers of responses. In addition to being substantively large and in the expected direction,the difference was statistically significant, with a t-ratio of 2.628 (p=.009). Again, the results are in line with other research showing improved response rates associated with personalized salutations.

Table 2.Hypothesis Test of E-mail Personalization, Cable Survey, 2007.

| | Personalized | Non-Personalized | Difference | Total |
|---|---|---|---|---|
| Respondent Count | 326 | 265 | 61 | 591 |
| Non-Respondent Count | 3139 | 3201 | -62 | 6340 |
| Total Count | 3465 | 3466 | -1 | 6931 |
| Response Rate (Mean) | 9.41% | 7.65% | 1.76% | 8.53% |
| Standard Deviation | 29.20 | 26.58 | | |

t = 2.628; df = 6929; sig (2-tailed) = 0.009

Experiments 4-6: The 2009 SUNY Student Opinion Survey (SOS)

Between March 18th and April 30th, 2009 The University at Albany surveyed its undergraduate student population on a variety of areas related to student satisfaction and their educational experiences as part of the SUNY-wide administration of the Student Opinion Survey (SOS), a survey effort going back to the 1980s. The surveys were conducted on UAlbany's behalf by American College Testing (ACT). Two days before the first invitation, we sent a pre-notification e-mail (which was the subject of these experiments) and then ACT sent out up to three e-mail invitations to all matriculated undergraduates requesting their participation. In addition, as with the SES in 2007, deans, department chairs, program directors and advisors were asked to send their students e-mails requesting their participation in the survey. As an incentive for participation, students who completed the survey were offered the chance to participate in a drawing for a single cash prize of $250.00.

Overall, 2,226 students participated in the survey, representing 18.7% of UAlbany's undergraduate populationof 12,122. After ACT removed partial and spoiled surveys[2] 1,952 students remained, representing 16.1 percent of the population. It is this group we will examine first and count as completed surveys.

Having previously demonstrated the effectiveness of use of a personalized salutation in e-mail invitations for two very different types of surveys, and both for a first invitation e-mail and a third and final follow-up e-mail, we were interested in determining whether sending a pre-notification would help with our response rates, and if so, whether personalization has a similar impact with the pre-notification as it does with an invitation or reminder e-mail. In addition, we were interested in testing whether it would make a difference if the source of the e-mail was actually from the Vice Provost's e-mail account, or from a generic "UAlbany Survey" e-mail account which our office had recently set up for use on surveys. In addition, while we were aware of literature on the efficacy of pre-notification e-mails for mixed-mode surveys discussed above, we also shared Couper's skepticism that e-mail pre-notifications would have the same impact for a web-based survey for which e-mail invitations were being sent out to the same e-mail account as the pre-notification. Thus, an additional control group was not sent a pre-notification at all.

All students were sent the same e-mail text, signed by the Vice Provost for Undergraduate Education, with the subject line "UAlbany Student Opinion Survey." E-mails sent from the "UASurvey" account included the text "The following message is being sent to you on behalf of [name], Vice Provost for Undergraduate Education" at the top; those sent directly from the Vice Provost's account did not include this. As previously, the solicitation was either "Dear [first name]" or "Dear UAlbany Student" depending on the group to which the student was randomly assigned.

[2] Roughly 2/3 of the way through the survey, a question asks respondents to select "NA" as a way of weeding out students who might have been simply checking boxes in order to get to the end and qualify for the drawing.

Given the total population size of over 12,000, it was not a problem to randomly divide students into a total of five total experimental treatments:

1) No pre-notification (2,122 students)
2) Non-personalized pre-notification sent from UASurvey account (2,000)
3) Non-personalized pre-notification sent from VP's e-mail account (2,000)
4) Personalized pre-notification sent from UASurvey account (2,000)
5) Personalized pre-notification sent from VP's e-mail account (2,000)

As shown in Tables 3a-3e, none of the experimental treatments produced response rates higher than the control group which received no pre-notification; in fact, the reverse was true – every experimental treatment group had a slightly *lower* response rate than the control group. What's more, the differences among the four experimental treatment groups were negligible. Overall, the control group had a response rate of 16.8%, while the four treatment groups had remarkably similar response rates ranging between 15.6% and 15.9%.

Table 3b, below, shows the comparison of the control group and all four pre-notification groups combined. Overall, the students who received pre-notifications had a 15.8% response rate, about a point lower than the 16.8% response rate for the control group, a modest difference, and one in the opposite of the expected direction. This difference was not statistically significant, with a t-ratio of 1.461 (p=.144).

Tables 3c and 3d show even smaller differences among the groups that received pre-notifications. Here, the personalized salutation had no impact at all. Similarly, using the Vice Provost's own e-mail account rather than the "UAlbany Survey" account made no difference at all. Finally, Table 3e confirms that we cannot reject the null hypothesis that no significant differences exist among any of the five groups – the control group and the four experimental groups. Within-group variance dwarfs between-group variance and the overall F-statistic does not even come close to statistical significance.

These null findings are consistent with Couper's skepticism mentioned earlier (p. 306) that an e-mail pre-notification for a web survey would be of any use. In fact, it may be that the reverse is true, if the additional, apparently pointless, e-mail sours some prospective respondents towards the survey. Based on this, I would suggest that any e-mail contact regarding a web-based survey should include a link to the survey or risk being counter-productive.

Table 3a. Survey Response Rate, by Pre-Notification Treatment, SOS 2009.

| | No Pre-Notification | Generic/ UASurvey | Generic/ VP | Personalized/ UASurvey | Personalized / VP | Total |
|---|---|---|---|---|---|---|
| Respondent Count | 692 | 317 | 314 | 312 | 317 | 1952 |
| Non-Respondent Count | 3430 | 1683 | 1686 | 1688 | 1683 | 10171 |
| Total Count | 4122 | 2000 | 2000 | 2000 | 2000 | 12122 |
| Response Rate | 16.79% | 15.85% | 15.70% | 15.60% | 15.85% | 16.10% |

Table 3b.Hypothesis Test of Pre-Notification Efficacy, SOS 2009.

| | No Pre-Notification | Pre-Notification (All Types) |
|---|---|---|
| Respondent Count | 692 | 1260 |
| Non-Respondent Count | 3430 | 6740 |
| Total Count | 4122 | 8000 |
| Response Rate (Mean) | 16.79% | 15.75% |
| Standard Deviation | 37.38 | 36.43 |

t = 1.461; df = 12120; sig (2-tailed) = 0.144

Table 3c.Hypothesis Test of Pre-Notification Personalization, SOS 2009.

| | Generic Salutation | Personalized Salutation |
|---|---|---|
| Respondent Count | 631 | 629 |
| Non-Respondent Count | 3369 | 3371 |
| Total Count | 4000 | 4000 |
| Response Rate (Mean) | 15.78% | 15.73% |
| Standard Deviation | 36.46 | 36.41 |

t = 0.061; df = 7998; sig (2-tailed) = 0.951

Table 3d.Hypothesis Test, VP E-mail Account, SOS 2009.

| | UASurvey Account | VP's E-mail Account |
|---|---|---|
| Respondent Count | 629 | 631 |
| Non-Respondent Count | 3371 | 3369 |
| Total Count | 4000 | 4000 |
| Response Rate (Mean) | 15.73% | 15.78 % |
| Standard Deviation | 36.41 | 36.46 |

t = 0.061; df = 7998; sig (2-tailed) = 0.951

Table 3e.Anova analysis summary comparing response rate within all 5 groups.

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | .302 | 4 | .076 | .559 | .693 |
| Within Groups | 1637.368 | 12117 | .135 | | |
| Total | 1637.670 | 12121 | | | |

## Summary and Conclusion

The main findings of the six experiments presented here include:

- Confirmation that personalized solicitations do significantly improve response rates;
- Content of the e-mail's subject line (at least the language included in our experiments) does not significantly affect response rates;
- While keeping the signed sender constant, the actual e-mail account from which a pre-notification e-mail was sent does not affect response rate;
- Regardless of personalization or e-mail account source, sending an e-mail pre-notification for a web-based survey did not increase response rates and may even have decreased them.

The finding regarding personalization of the e-mail solicitation seems to be very robust in a variety of conditions, and confirms a growing body of existing data. I believe it is safe to say that this should be considered a best practice of web survey research among student populations, as it has already been for mail survey research for some time. Having said that, I should point out that the nature of the ideal salutation will of necessity be dependent on the nature of the population (see, e.g., Dillman, et al., p. 272). For the surveys of university undergraduates discussed here, first name seems to be an effective salutation; that may not be the case with faculty or other more professional populations.[3]

The finding regarding lack of utility of pre-notification e-mails is also quite strong and indicates that use of pre-notifications of this type is probably at best a waste of time, and at worst may turn some students off.

I should also note that these and other findings described here may or may not be applicable to other types of populations, and in fact, may not even be applicable across a variety of university settings, where typical response rates among students varies wildly from one campus to another.

The next step in this research is to examine carefully whether any of the experimental groups discussed in this paper differ significantly or substantively (1) with regard to either the demographic and academic characteristics of their populations, (2) or with regard to the substantive responses to the survey. Ideally, I would like to be able to do split-sample experiments regarding the use of lottery-style incentives, and their nature (e.g., use vs. non-use; use of one larger vs. several smaller prizes), but because we are conducting our surveys within a fairly small and self-contained population, it would not be advisable to create financial disparities in how our students are treated. However, I do plan on conducting experiments regarding how the incentive is described in the e-mail or e-mail subject line.

---

[3] In fact, when in the past I have used personalized salutations for faculty surveys, I found that this led to uncertainty over choosing the appropriate salutation (first name, full name, job title, etc.) along with raising suspicions among the faculty that confidentiality or anonymity would not be protected. While the latter is just anecdotal, I have concluded that a simple "Dear Colleague" salutation is probably better for surveys of faculty and staff.

## II. Worth the Weight? The Benefits and Pitfalls of

## Applying Post-Stratification Weights to Web Surveys of College Undergraduates

Introduction

In most fields of survey research it is customary to weight respondent data to known population parameters when it is observable that they differ due to differential selection probabilities or nonresponse bias. As Lewis Mandell describes the problem,

> Upon completion in a sample survey, the researcher often finds that the response rate is not uniform across all subgroups; rather there are differences among various segments of the population. This, in itself, introduces no bias in population estimates since it is theoretically possible that responses are similar for subgroups with varying response rates. In actual practice, however, the conditions determining the probability of response are also likely to affect responses.
>
> In this manner, differential nonresponse may introduce bias in population estimates. (Mandell, 1974)

Thus, the main reason to weight the data is to improve survey estimates, in case there are important differences in response patterns between over-represented and under-represented sub-populations. Another reason for weighting is to make the survey more fully representative of the population from which it is drawn, for instances in which that might be an important goal in and of itself, either for reasons of equity or political considerations. This type of weighting can be done easily for any characteristics for which population parameters are known.

One such domain in which a great many surveys are conducted and in which population parameters are well known is within a college or university. Colleges throughout the U.S. and Canada (as well as elsewhere) regularly survey their students and other populations on a variety of topics, most importantly on self-assessments of their academic experiences, engagement and satisfaction.

Yet, perhaps because most academic administrators are not trained statisticians one rarely hears requests for weighted data of these surveys. Administrators want to know the "survey results" or "what the survey says," but do not generally request analysis of weighted data. To the contrary, among administrators and representatives of faculty governance, any post-survey weighting schemes may even be viewed incorrectly as tampering with the survey data.

Yet surveys are used for assessment purposes, including accreditation, making accurate estimates particularly important, especially when estimates from more than one survey are compared over time. For multi-institutional surveys, institutions are often compared with one another with little or no attention to ways in which their samples might differ in

non-random ways. In some instances, these cross-institutional comparisons are even made when the surveys are conducted using entirely different modes of administration. It was with one particular such survey in mind that I began to think about the differences that weighting might make in surveys of student populations.

## Data and Analysis

For this portion of the paper I analyze data from three surveys, summarized below in Table 4. The 2006 Student Opinion Survey (SOS) was administered using scannable paper forms in a sample of undergraduate classes at the University at Albany, SUNY (UAlbany) between March 30th and April 4th, 2006. Total enrollment in the sampled classes was 928, of whom 645 students (70%) were present the days the survey was administered. A total of 597 students participated in the survey, yielding 583 useable surveys.[4] Therefore the cooperation rate was 93% and the response rate was 90%. Viewed as a percentage of the entire enrollment of the classes sampled, the cooperation rate was 64% and the response rate was 63%.

Because of the mode of administration, this survey potentially includes two types of bias – first, bias due to differential probability of selection based on the classes sampled, and second, due to nonresponse. The latter would be seen here more with regard to the 283 students who did not attend class on the day the survey was administered than the 48 who choose not to participate or the 14 who were excluded (see the footnote below).

Surveys with identical question wording and order were administered at roughly the same time throughout the State University of New York (SUNY) system, and comparisons were made among schools. In this case, both ordinal rankings and tests of statistical significance were conducted between UAlbany and both the other three SUNY university centers and all 26 state-operated 4-year colleges and universities throughout the system.

These comparisons were made despite the fact that different institutions administered their surveys in dramatically different ways. For our purposes here, the most important point is that of the four university centers, two (including UAlbany) administered their surveys by paper to a sample of classes and the other two administered theirs to all enrolled undergraduates using a web survey.

It is for this reason that I chose to analyze the Spring, 2007 UAlbany Student Experience Survey (SES). The Spring 2007 SES was administered to matriculated undergraduates via the internet between March 19th and May 11th, 2007. A total of 2,023 students, representing 18% of matriculated undergraduates, participated in the survey.

[4] Fourteen completed surveys were not included in the final sample because the respondents incorrectly answered a question designed to catch students who were just filling the surveys in down the line, without paying attention to the questions.

Thus we have one survey administered on paper to a small sample with a high response rate and another one administered by the internet to the full undergraduate population, with a low response rate but yielding a large sample. The surveys also differed with regard to their content. The 2006 SOS questions largely deal with student satisfaction while the 2007 SES questions largely deal with engagement and educational outcomes.

One central motivation for conducting the analysis below was that, with the Spring 2009 administration of the SOS coming up, I wanted to determine whether shifting our mode of administration from in-class to web-based would have any impact on the survey results. As I will show below, the analysis of the 2006 SOS and 2007 SES showed that differences in the mode of administration would not be likely to have an impact on the result, so we went ahead with web-based administration for the 2009 SOS. Thus, the final set of analysis is on that data set, with 1,952 valid responses received between March 17th and April 30th, 2009.

**Table 4: Summary of the Three Surveys**

| Survey Characteristics | SOS (2006) Student Opinion Survey | SES (2007) Student Experience Survey | SOS (2009) Student Opinion Survey |
|---|---|---|---|
| **Sampling** | Classroom | Population | Population |
| **Mode** | Scannable Paper | Web | Web |
| **Invitation** | In-Class | E-mail Invitations, Flyers | E-mail Invitations, Flyers |
| **"Sample" Size** | 583 | 2,023 | 1,952 |
| **Response Rate** | 63% | 18% | 17% |
| **Incentive** | None | 3 $50 Prizes | 1 $250 Prize |
| **Content** | Student Satisfaction | Student Activities, Learning Outcomes | Student Satisfaction |
| **Uses** | Time Series, Benchmarks w/ SUNY | Time Series, Outcome-Based Assessment | Time Series, Benchmarks w/ SUNY |

The 2006 Student Opinion Survey

Table 5, below, shows sample and population demographics for four variables: ethnicity, gender, student level (freshman through senior) and admission type (freshman vs. transfer). These are not meant by any means to be a comprehensive list of variables by which we might consider weighting; rather, they are meant to represent several variables that we might expect to have important impacts on response patterns, and that are also matters of critical interest to university administrators. The reduced sample size of 519 is due to the fact that 64 survey instruments did not include a useable student identification number, which was needed in order to match survey data to the student data file.

**Table 5: UAlbany 2006 Student Opinion Survey: Sample and Population Demographics**

| Race/Ethnicity | Frequency | Sample Percent | Population Percent | Difference | Prelim. Weight | Final Weight |
|---|---|---|---|---|---|---|
| White | 372 | 71.7 | 60.0 | 11.7 | 0.84 | 0.84 |
| Black | 31 | 6.0 | 8.3 | -2.3 | 1.38 | 1.38 |
| Hispanic | 23 | 4.4 | 7.3 | -2.9 | 1.66 | 1.66 |
| Asian or Pacific Islander | 21 | 4.0 | 5.6 | -1.6 | 1.40 | 1.40 |
| Amer. Indian or Alaska Nat. | 1 | 0.2 | 0.3 | -0.1 | NA | NA |
| Non-Resident | 8 | 1.5 | 1.9 | -0.4 | NA | NA |
| Unknown | 63 | 12.1 | 16.7 | -4.6 | 1.38 | 1.38 |
| Total | 519 | 100.0 | 100.0 | | | |
| Sex/Gender | Frequency | Sample Percent | Population Percent | Difference | Prelim. Weight | Final Weight |
| Female | 273 | 52.6 | 50.5 | 2.1 | 0.96 | 0.93 |
| Male | 246 | 47.4 | 49.5 | -2.1 | 1.04 | 1.07 |
| Total | 519 | 100.0 | 100.0 | | | |
| Student Level | Frequency | Sample Percent | Population Percent | Difference | Prelim. Weight | Final Weight |
| Freshman | 124 | 23.9 | 17.7 | 6.2 | 0.74 | 0.74 |
| Sophomore | 113 | 21.8 | 22.3 | -0.5 | 1.02 | 1.02 |
| Junior | 165 | 31.8 | 29.2 | 2.6 | 0.92 | 0.92 |
| Senior | 117 | 22.5 | 30.8 | -8.3 | 1.37 | 1.37 |
| Total | 519 | 100.0 | 100.0 | | | |
| Admission Type | Frequency | Sample Percent | Population Percent | Difference | Prelim. Weight | Final Weight |
| Freshman | 346 | 66.7 | 65.4 | 1.3 | NA | NA |
| Transfer | 173 | 33.3 | 34.6 | -1.3 | NA | NA |
| Total | 519 | 100.0 | 100.0 | | | |

Starting at race and ethnicity, the largest difference we see is that whites comprise 72% of the sample but only 60% of the population. Other groups, including Blacks, Hispanics, Asians, and "unknowns" are all under-represented.[5] With regard to gender, women are slightly overrepresented and men slightly underrepresented. Looking at student level, freshmen are overrepresented[6] and seniors are underrepresented, with sophomores and juniors coming closer to population parameters.

Table 6, below, shows results for four selected survey questions that get at overall satisfaction, both for the whole sample and cross-tabulated by the demographic groups discussed above. To facilitate interpretation, I have highlighted cells that have fairly large differences among groups (highlighting does not necessarily indicate statistical significance. Without getting into the details of the individual survey times, we see first of all that gender does not seem to have had much impact on response patterns for these questions. The only question that shows any substantial difference is the one asking whether they would choose UAlbany again if they had it to do over. Using a scale of 1 to 5, the average response was higher for men than for women, indicating that male students were more likely to feel that they made the right choice.

On race and ethnicity, we see larger differences that operate systematically across all four questions. First of all, Hispanic or Latino students in the sample responded substantially more positively on all four items. On the other hand, Asian Americans responded substantially more negatively. African Americans had more mixed responses – roughly the same as the overall population on one item, more negative on two and more positive on one.

Looking at student level, class rank does not seem to be an important correlate with any of the selected survey items – differences among classes for all items are small. Finally, transfer students had slightly more positive evaluations of UAlbany than freshman admits across all four survey items selected.

To summarize what we have seen so far, the survey sample deviated substantially from population parameters in two of the four demographic categories – race and student level. As shown in Table 5, the sample deviated by a modest amount with regard to gender. Finally, with regard to admission type, the survey sample deviated only by around one percentage point. As shown in Table 6, response patterns differed substantially only by race and ethnicity, and only slightly by the other factors.

There is no critical test to determine whether to weight by a particular variable or combination of variables. Given the combination of demographic properties of the sample and response patterns on the survey item, the order of importance for weighting would clearly place race first. Just as clearly, admission type would be last, with sex and student level in between. Under these circumstances it would be justified to weight only by race/ethnicity, but for purposes of this paper as an academic exercise, I have chosen to weight by sex and student level as well.

---

[5] For purposes of this paper, I use the SUNY system's names for racial and ethnic categories, simply because those are the categories that exist in our student data records.

[6] The reason for the apparently low percentage of freshmen in the population is that this variable is determined by total credits, including transfer and AP credits. Thus, in the Spring semester, many students appear to move up a class.

**Table 6: Responses to Selected SOS Questions by Demographic Categories.**

| # | Question/Response | Total | Female | Male | White | Black | Hispanic | Asian | Unknown | Frosh | Soph | Junior | Senior | FrAdmit | TrAdmit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | n=519 | n=273 | n=246 | n=372 | n=31 | n=23 | n=21 | n=63 | n=124 | n=113 | n=164 | n=117 | n=345 | n=173 |
| **Si1** | **Academic experiences have:** | | | | | | | | | | | | | | |
| | Not met expectations (1) | 13.5 | 13.6 | 13.4 | 13.7 | 6.7 | 4.3 | 28.6 | 14.3 | 11.3 | 13.3 | 14.6 | 14.5 | 16.8 | 6.9 |
| | Met expectations (2) | 73.0 | 72.4 | 74.0 | 71.8 | 83.3 | 69.6 | 71.4 | 76.2 | 75.8 | 72.6 | 72.6 | 71.8 | 69.9 | 79.8 |
| | Exceeded expectations (3) | 13.3 | 14.0 | 12.6 | 14.5 | 10.0 | 26.1 | 0.0 | 9.5 | 12.9 | 14.2 | 12.8 | 13.7 | 13.3 | 13.3 |
| | Average | 2.00 | 2.00 | 1.99 | 2.01 | 2.03 | 2.22 | 1.71 | 1.95 | 2.02 | 2.01 | 1.98 | 1.99 | 1.97 | 2.06 |
| **Si3** | **Would choose UAlbany again:** | | | | | | | | | | | | | | |
| | Definitely No (1) | 5.2 | 5.9 | 4.5 | 4.9 | 6.7 | 0.0 | 4.8 | 7.9 | 4.9 | 7.1 | 5.5 | 3.4 | 4.9 | 5.8 |
| | Probably No (2) | 9.6 | 11.0 | 8.2 | 10.5 | 13.3 | 0.0 | 9.5 | 7.9 | 8.9 | 8.0 | 9.8 | 12.0 | 11.0 | 6.9 |
| | Uncertain (3) | 17.2 | 18.8 | 15.5 | 18.3 | 16.7 | 4.3 | 28.6 | 11.1 | 20.3 | 18.6 | 14.0 | 17.1 | 18.3 | 15.0 |
| | Probably Yes (4) | 39.1 | 36.8 | 41.6 | 37.7 | 40.0 | 43.5 | 38.1 | 42.9 | 39.0 | 32.7 | 40.9 | 42.7 | 38.4 | 40.5 |
| | Definitely Yes (5) | 28.8 | 27.6 | 30.2 | 28.6 | 23.2 | 52.2 | 19.0 | 30.2 | 26.8 | 33.6 | 29.9 | 24.8 | 27.3 | 31.8 |
| | Average | 3.77 | 3.69 | 3.85 | 3.75 | 3.60 | 4.48 | 3.57 | 3.79 | 3.74 | 3.78 | 3.80 | 3.74 | 3.72 | 3.86 |
| **Si6** | **Quality of Education is:** | | | | | | | | | | | | | | |
| | Very Low (1) | 0.6 | 0.4 | 0.8 | 0.5 | 0.0 | 0.0 | 4.8 | 0.0 | 2.4 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 |
| | Low (2) | 1.9 | 2.2 | 1.6 | 1.6 | 0.0 | 0.0 | 0.0 | 6.5 | 0.0 | 1.8 | 3.7 | 1.7 | 2.6 | 0.6 |
| | Average (3) | 42.2 | 43.5 | 40.7 | 41.5 | 32.3 | 26.1 | 52.4 | 48.4 | 42.7 | 40.7 | 38.4 | 48.3 | 42.9 | 40.7 |
| | High (4) | 49.1 | 46.9 | 51.6 | 50.4 | 61.3 | 60.9 | 42.9 | 37.4 | 46.8 | 53.1 | 52.4 | 43.1 | 49.0 | 49.4 |
| | Very High (5) | 6.2 | 7.0 | 5.3 | 5.9 | 6.5 | 13.0 | 0.0 | 8.1 | 8.1 | 4.4 | 5.5 | 6.9 | 4.6 | 9.3 |
| | Average | 3.58 | 3.58 | 3.59 | 3.60 | 3.74 | 3.87 | 3.33 | 3.47 | 3.58 | 3.60 | 3.60 | 3.55 | 3.54 | 3.67 |
| **Si7** | **Overall Satisfaction:** | | | | | | | | | | | | | | |
| | Very Dissatisfied (1) | 0.8 | 0.7 | 0.8 | 0.5 | 0.0 | 0.0 | 4.8 | 1.6 | 1.6 | 0.0 | 0.6 | 0.9 | 0.9 | 0.6 |
| | Dissatisfied (2) | 6.6 | 6.6 | 6.6 | 5.9 | 9.7 | 0.0 | 14.3 | 7.9 | 4.9 | 6.2 | 6.1 | 9.4 | 7.0 | 5.8 |
| | Neither Sat. nor Diss. (3) | 16.1 | 16.1 | 16.0 | 18.6 | 12.9 | 0.0 | 4.8 | 12.7 | 16.3 | 18.6 | 14.6 | 15.4 | 18.0 | 12.2 |
| | Satisfied (4) | 61.3 | 61.9 | 60.7 | 59.7 | 71.0 | 65.2 | 76.2 | 57.1 | 62.6 | 58.4 | 64.0 | 59.0 | 60.0 | 64.0 |
| | Very Satisfied (5) | 15.3 | 14.7 | 16.0 | 15.1 | 6.5 | 34.8 | 0.0 | 20.6 | 14.6 | 16.8 | 14.6 | 15.4 | 14.2 | 17.4 |
| | Average | 3.84 | 3.83 | 3.84 | 3.83 | 3.74 | 4.35 | 3.52 | 3.87 | 3.84 | 3.86 | 3.86 | 3.79 | 3.80 | 3.92 |

The last two columns of Table 5, above, show the preliminary weight for each demographic category. This is simply the population percentage divided by the sample percentage (see, e.g., Groves et al., 2004, p. 326). For under-represented groups, this figure will thus be greater than "1" and for over-represented groups it will be less than "1." The total weight variable is simply the product of all the individual weight variables (Groves, 2004; Mandell, 1974).

Because of differentials in the ways in which each group is represented in interaction with the others, this initial round of weighting generally does not produce "perfect" matches to population parameters, requiring a few rounds of iterative tweaking to the weights. The final column of Table 5 shows the final weights used for this analysis. Finally, Table 7, below, shows that the weighting procedure has gotten us a great deal closer to the population parameters. While it is likely that additional tweaking could get us even closer, these distributions are well within standard sampling error protocols.

**Table 7: UAlbany 2006 Student Opinion Survey: Weighted Demographics**

| Race/Ethnicity | Unweighted Frequency | Sample Percent | Weighted Frequency | Sample Percent | Population Percent | Difference |
|---|---|---|---|---|---|---|
| White | 372 | 71.7 | 309 | 59.6 | 60.0 | -0.4 |
| Black | 31 | 6.0 | 41 | 8.0 | 8.3 | -0.3 |
| Hispanic | 23 | 4.4 | 37 | 7.2 | 7.3 | -0.1 |
| Asian or Pacific Islander | 21 | 4.0 | 31 | 6.0 | 5.6 | 0.4 |
| Amer. Indian or Alaska Nat. | 1 | 0.2 | 1 | 0.2 | 0.3 | -0.1 |
| Non-Resident | 8 | 1.5 | 8 | 1.6 | 1.9 | -0.3 |
| Unknown | 63 | 12.1 | 90 | 17.4 | 16.7 | 0.7 |
| Total | 519 | 100.0 | 518 | 100.0 | 100.0 | |
| **Sex/Gender** | **Unweighted Frequency** | **Sample Percent** | **Weighted Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| FEMALE | 273 | 52.6 | 262 | 50.6 | 50.5 | 0.1 |
| MALE | 246 | 47.4 | 256 | 49.4 | 49.5 | -0.1 |
| Total | 519 | 100.0 | 518 | 100.0 | 100.0 | |
| **Student Level** | **Unweighted Frequency** | **Sample Percent** | **Weighted Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 124 | 23.9 | 90 | 17.4 | 17.7 | -0.3 |
| Sophomore | 113 | 21.8 | 112 | 21.7 | 22.3 | -0.6 |
| Junior | 165 | 31.8 | 151 | 29.2 | 29.2 | 0.0 |
| Senior | 117 | 22.5 | 165 | 31.8 | 30.8 | 1.0 |
| Total | 519 | 100.0 | 518 | 100.0 | 100.0 | |
| **Admission Type** | **Unweighted Frequency** | **Sample Percent** | **Weighted Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 346 | 66.7 | 339 | 65.4 | 65.4 | 0.0 |
| Transfer | 173 | 33.3 | 179 | 34.6 | 34.6 | 0.0 |
| Total | 519 | 100.0 | 518 | 100.0 | 100.0 | |
| Weight Variable Minimum: | 0.58 | | | | | |
| Weight Variable Maximum: | 2.43 | | | | | |

Due to the small sample size for non-white racial groups, I was unable to conduct a more sophisticated weighting that takes into account differential response rates and response distributions by race and gender combined. Weighting simply by broad groups without cross-tabulation requires an assumption that may or may not be merited here: "that within subgroups…the respondents are a random sample of all sample persons" (Groves, 2004). I will discuss this matter in more detail in the analysis of the 2007 Student Experience Survey, with its larger sample that enables that level of analysis.

The final question here is whether the weighting has made any difference in the survey results. As shown in Table 8, below, the answer is clearly, "no, it has not." Whether looking at percentages of individual response options, the combined top two most positive responses, or average response, the differences are miniscule.

**Table 8: UAlbany 2006 SOS: Weighted Survey Results**

| # | Question/Response | Unweighted n=519 | Weighted n=518 |
|---|---|---|---|
| **Si1** | **Academic experiences have:** | | |
| | Not met expectations (1) | 13.5 | 13.5 |
| | Met expectations (2) | 73.0 | 73.4 |
| | Exceeded expectations (3) | 13.3 | 13.1 |
| | Average | 2.00 | 2.00 |
| **Si3** | **Would choose UAlbany again:** | | |
| | Definitely No (1) | 5.2 | 5.0 |
| | Probably No (2) | 9.6 | 9.4 |
| | Uncertain (3) | 17.2 | 16.5 |
| | Probably Yes (4) | 39.1 | 39.8 |
| | Definitely Yes (5) | 28.8 | 29.3 |
| | Top Two Categories | 67.9 | 69.1 |
| | Average | 3.77 | 3.79 |
| **Si6** | **Quality of Education is:** | | |
| | Very Low (1) | 0.6 | 0.4 |
| | Low (2) | 1.9 | 2.1 |
| | Average (3) | 42.2 | 42.5 |
| | High (4) | 49.1 | 48.8 |
| | Very High (5) | 6.2 | 6.2 |
| | Top Two Categories | 55.3 | 55.0 |
| | Average | 3.58 | 3.58 |
| **Si7** | **Overall Satisfaction:** | | |
| | Very Dissatisfied (1) | 0.8 | 0.8 |
| | Dissatisfied (2) | 6.6 | 7.0 |
| | Neither Sat. nor Diss. (3) | 16.1 | 15.0 |
| | Satisfied (4) | 61.3 | 61.3 |
| | Very Satisfied (5) | 15.3 | 16.0 |
| | Top Two Categories | 76.6 | 77.3 |
| | Average | 3.84 | 3.85 |

The 2007 Student Experience Survey

Because the 2007 Student Experience Survey (SES) was conducted online, with student identification numbers used for login, we were able to match all 2,023 cases to data in the student data file. As shown in Table 5, below, the web administration resulted in a very different demographic distribution than the in-class sample survey used a year earlier for the SOS. While the SOS greatly over-represented white students, the SES did so by a smaller amount. On the other hand, the SES sample still under-represented Blacks, Hispanics and Asian Americans. The biggest difference between the two samples[7] is gender – while the SOS sample slightly over-represented women, the SES sample did so by a very large amount. While the population was 49% female, the sample was 63% female. The SES sample was much more representative than the SOS sample with regard to student level, with only small differences observed. However, unlike the SOS, the SES sample substantially over-represented freshman admits at the expense of transfers.

For our purposes, the most important difference between the SOS and SES surveys is that the latter has a sample of over 2,000, meaning that we can do a much more fine-tuned job of weighting by cross-tabulated subgroups. As we discussed earlier, simply weighting separately by two factors necessitates the assumption that the response patterns between those two factors are not correlated. This is called the "missing at random" assumption (Groves, 2004). While we have good reason to make that assumption with regard to the other factors (student level and admit type), we know for sure that the missing at random assumption does not apply with regard to race and gender.

As Table 9 shows, in addition to a relationship between race or gender and response rate, response rate (shown here in terms of the degree to which a sub-group is over- or under-represented compared to the population) varies within race and gender categories as well. So African American women are only slightly under-represented (5.1% of the sample compared to 5.4% of the population) while African American men are tremendously under-represented (1.1% of the sample compared to 3.3% of the population). Hispanic/Latina women are not under-represented, while Hispanic/Latino men are seriously under-represented. So weighting by both race and gender seems to be indicated for this survey. This is accomplished, as shown in Table 6, simply by subdividing the sample one additional degree (in this case by race and gender) and creating a weight variable with separate values for each of the now subdivided cells.

The final column of Table 9 shows these calculated weight factors for race, subdivided by gender, and for the other factors by themselves. In cases of cell sizes below 20 cases, the overall weight for the gender was used instead of the calculated weight for the subgroup with the small cell size (see Native American women and non-resident men). In addition to being more statistically valid (due to the fact that we need not rely on the missing at random assumption), this method also has the advantage of being more fine-tuned, reducing the need for additional iterations and fine-tuning of the weights. In this case, after the initial round of weighting, no additional re-weighting was required, making the initial weights also the final weights.

[7] For purposes of convenience, I will refer to the group of students who chose to take the Student Experience Survey as a sample, even though the entire undergraduate student body was invited to participate in the survey, meaning that no sampling was actually involved.

**Table 9: UAlbany 2007 SES: Sample and Population Demographics**

| Female | Frequency | Sample Percent | Population Percent | Difference | Weight |
|---|---|---|---|---|---|
| White | 771 | 38.1% | 28.5% | 9.6% | 0.75 |
| Black | 103 | 5.1% | 5.4% | -0.3% | 1.05 |
| Hispanic | 82 | 4.1% | 4.1% | 0.0% | 1.00 |
| Asian or Pacific Islander | 78 | 3.9% | 2.7% | 1.2% | 0.70 |
| Amer. Indian or Alaska Nat. | 3 | 0.1% | 0.1% | 0.0% | 0.78 |
| Non-Resident | 24 | 1.2% | 0.9% | 0.3% | 0.72 |
| Unknown | 212 | 10.5% | 7.8% | 2.7% | 0.74 |
| Total, Female | 1273 | 62.9% | 49.4% | 13.6% | 0.78 |
| **Male** | | | | | |
| White | 495 | 24.5% | 30.9% | -6.4% | 1.26 |
| Black | 23 | 1.1% | 3.3% | -2.2% | 2.94 |
| Hispanic | 41 | 2.0% | 3.5% | -1.4% | 1.71 |
| Asian or Pacific Islander | 46 | 2.3% | 2.6% | -0.3% | 1.15 |
| Amer. Indian or Alaska Nat. | 0 | 0.0% | 0.1% | -0.1% | NA |
| Non-Resident | 7 | 0.3% | 1.0% | -0.6% | 1.37 |
| Unknown | 138 | 6.8% | 9.3% | -2.4% | 1.36 |
| Total, Male | 750 | 37.1% | 50.6% | -13.6% | 1.37 |
| Total, Sample | 2023 | 100.0% | 100.0% | 0.0% | |

| Student Level | Frequency | Sample Percent | Population Percent | Difference | Weight |
|---|---|---|---|---|---|
| Freshman | 365 | 18.0 | 17.1 | 0.9 | 0.95 |
| Sophomore | 497 | 24.6 | 24.9 | -0.3 | 1.00 |
| Junior | 525 | 26.0 | 28.2 | -2.2 | 1.09 |
| Senior | 636 | 31.4 | 29.9 | 1.6 | 0.95 |
| Total | 2023 | 100.0 | 100.0 | | |

| Admission Type | Frequency | Sample Percent | Population Percent | Difference | Weight |
|---|---|---|---|---|---|
| Freshman | 1431 | 70.7 | 65.0 | 5.7 | 0.92 |
| Transfer | 590 | 29.2 | 35.0 | -5.8 | 1.20 |
| Total | 2021 | 100.0 | 100.0 | | |

Table 10, below, shows the weighted frequencies and percentages for the 2007 Student Experience Survey.The final three columns show that in no case was the weighted percentage for the group or subgroup off by more than one half of a percentage point from the population value.

**Table 10: UAlbany 2007 SES: Weighted Demographics**

| Female | *Unweighted* Frequency | Sample Percent | *Weighted* Frequency | Sample Percent | Population Percent | Difference |
|---|---|---|---|---|---|---|
| White | 771 | 38.1% | 575 | 28.8% | 28.5% | 0.3% |
| Black | 103 | 5.1% | 104 | 5.2% | 5.4% | -0.2% |
| Hispanic | 82 | 4.1% | 79 | 4.0% | 4.1% | -0.1% |
| Asian or Pacific Islander | 78 | 3.9% | 53 | 2.7% | 2.7% | 0.0% |
| Amer. Indian or Alaska Nat. | 3 | 0.1% | 3 | 0.2% | 0.1% | 0.1% |
| Non-Resident | 24 | 1.2% | 18 | 0.9% | 0.9% | 0.0% |
| Unknown | 212 | 10.5% | 155 | 7.8% | 7.8% | 0.0% |
| Total, Female | 1273 | 62.9% | 987 | 49.4% | 49.4% | 0.0% |
| **Male** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| White | 495 | 24.5% | 621 | 31.1% | 30.9% | 0.2% |
| Black | 23 | 1.1% | 67 | 3.4% | 3.3% | 0.1% |
| Hispanic | 41 | 2.0% | 69 | 3.5% | 3.5% | 0.0% |
| Asian or Pacific Islander | 46 | 2.3% | 51 | 2.6% | 2.6% | 0.0% |
| Amer. Indian or Alaska Nat. | 0 | 0.0% | 0 | 0.0% | 0.1% | -0.1% |
| Non-Resident | 7 | 0.3% | 10 | 0.5% | 1.0% | -0.5% |
| Unknown | 138 | 6.8% | 191 | 9.6% | 9.3% | 0.3% |
| Total, Male | 750 | 37.1% | 1009 | 50.6% | 50.6% | 0.0% |
| Total, Sample | 2023 | 100.0% | 1996 | 100.0% | 100.0% | 0.0% |
| **Student Level** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 365 | 18.0 | 339 | 17.0% | 17.1% | -0.1% |
| Sophomore | 497 | 24.6 | 487 | 24.4% | 24.9% | -0.5% |
| Junior | 525 | 26.0 | 565 | 28.3% | 28.2% | 0.1% |
| Senior | 636 | 31.4 | 606 | 30.4% | 29.9% | 0.5% |
| Total | 2023 | 100.0 | 1998 | 100.0% | 100.0% | |
| **Admission Type** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 1431 | 70.7 | 1295 | 64.8% | 65.0% | -0.2% |
| Transfer | 590 | 29.2 | 702 | 35.2% | 35.0% | 0.2% |
| Total | 2021 | 100.0 | 1998 | 100.0% | 100.0% | |
| Weight Variable Minimum: | 0.61 | | | | | |
| Weight Variable Maximum: | 3.74 | | | | | |

Finally, we get to the question as to whether the weighting affected the survey results. As mentioned earlier, the SES does not contain the same type of satisfaction questions that the SOS has. As a result I chose the single item that deals with general satisfaction and then chose three other items that are of particular interest to academic administrators at UAlbany and presumably elsewhere: contribution to writing effectively, contribution to evaluating ideas critically, and whether students want more from academic advisement than they currently receive.

As with the SOS, the weighting has had no discernable impact. One item (advisement) is slightly more positive after the weighting, while the other three items are slightly more negative.

**Table 11: UAlbany 2007 SES: Weighted Survey Results**

| # | Question/Response | Unweighted | Weighted | |
|---|---|---|---|---|
| | | n=2023 | n=1998 | Difference |
| **s3q1** | **Satisfied with Academic Experiences:** | | | |
| | Never (1) | 1.1 | 1.2 | 0.1 |
| | Rarely (2) | 8.0 | 8.2 | 0.2 |
| | Sometimes (3) | 33.1 | 33.2 | 0.1 |
| | More Often Than Not (4) | 45.6 | 45.6 | 0.0 |
| | Almost Always (5) | 12.2 | 11.5 | -0.7 |
| | Top 2 Categories | 57.8 | 57.1 | -0.7 |
| | Average | 3.60 | 3.58 | -0.02 |
| **Si3** | **UAlbany's Contribution to Writing Effectively** | | | |
| | None (1) | 8.9 | 9.2 | 0.3 |
| | Small (2) | 21.6 | 21.6 | 0.0 |
| | Moderate (3) | 39.1 | 39.5 | 0.4 |
| | Large (4) | 21.3 | 20.8 | -0.5 |
| | Very Large (5) | 8.7 | 8.4 | -0.3 |
| | Top 2 Categories | 30.0 | 29.2 | -0.8 |
| | Average | 2.99 | 2.97 | -0.02 |
| **Si6** | **UAlbany's Contribution to Evaluating Ideas Critically** | | | |
| | None (1) | 1.9 | 2.1 | 0.2 |
| | Small (2) | 8.6 | 8.9 | 0.3 |
| | Moderate (3) | 36.3 | 36.2 | -0.1 |
| | Large (4) | 39.5 | 39.6 | 0.1 |
| | Very Large (5) | 13.7 | 13.2 | -0.5 |
| | Top 2 Categories | 53.1 | 52.8 | -0.3 |
| | Average | 3.55 | 3.53 | -0.02 |
| **Si7** | **Want More From Advisement?** | | | |
| | Yes (1) | 34.5 | 34.1 | -0.4 |
| | No (2) | 65.5 | 65.9 | 0.4 |

Mode of Administration Effect? A Perverse Exercise

As mentioned earlier, the Student Opinion Survey (SOS) is used to compare institutions within the SUNY system, despite the fact that different schools use different modes of administration. In 2006, UAlbany administered the survey to a sample of classes; two of our three direct comparator schools (of the four total comprehensive university centers) used web administration to their entire undergraduate population – just as we did the next year for the Student Experience Survey (SES).

Having noted the large demographic differences in the make-up of the two samples, an additional question was raised: did our mode of administration hurt (or help) us in comparison with our peers? While that question cannot be answered directly without a true experiment, I thought it might be an interesting and worthwhile (if somewhat perverse) exercise to see what would happen if we weight the SOS survey results not to the population demographics, but rather to the SES sample demographics. As mentioned in the introduction, another reason for conducting this analysis was to determine whether we could shift our administration to the web for the 2009 SOS.

**Table 12: UAlbany 2006 SOS and SES Demographics.**

| Race/Ethnicity | Frequency | Sample Percent | SES Percent | Difference | Prelim. Weight | Final Weight |
|---|---|---|---|---|---|---|
| White | 372 | 71.7 | 62.6 | 9.1 | 0.87 | 0.89 |
| Black | 31 | 6.0 | 6.2 | -0.2 | 1.04 | 1.04 |
| Hispanic | 23 | 4.4 | 6.1 | -1.7 | 1.38 | 1.38 |
| Asian or Pacific Islander | 21 | 4.0 | 6.1 | -2.1 | 1.53 | 1.40 |
| Amer. Indian or Alaska Nat. | 1 | 0.2 | 0.1 | 0.1 | 0.74 | NA |
| Non-Resident | 8 | 1.5 | 1.5 | 0.0 | 1.00 | NA |
| Unknown | 63 | 12.1 | 17.3 | -5.2 | 1.43 | 1.30 |
| Total | 519 | 100.0 | 100.0 | | | |
| **Sex/Gender** | **Frequency** | **Sample Percent** | **SES Percent** | **Difference** | **Prelim. Weight** | **Final Weight** |
| Female | 273 | 52.6 | 62.9 | -10.3 | 1.20 | 0.80 |
| Male | 246 | 47.4 | 37.1 | 10.3 | 0.78 | 1.16 |
| Total | 519 | 100.0 | 100.0 | | | |
| **Student Level** | **Frequency** | **Sample Percent** | **SES Percent** | **Difference** | **Prelim. Weight** | **Final Weight** |
| Freshman | 124 | 23.9 | 18.0 | 5.9 | 0.75 | 0.77 |
| Sophomore | 113 | 21.8 | 24.6 | -2.8 | 1.13 | 1.13 |
| Junior | 165 | 31.8 | 26.0 | 5.8 | 0.82 | 0.82 |
| Senior | 117 | 22.5 | 31.4 | -8.9 | 1.40 | 1.40 |
| Total | 519 | 100.0 | 100.0 | | | |
| **Admission Type** | **Frequency** | **Sample Percent** | **SES Percent** | **Difference** | **Prelim. Weight** | **Final Weight** |
| Freshman | 346 | 66.7 | 70.9 | -4.2 | 1.1 | 1.1 |
| Transfer | 173 | 33.3 | 29.2 | 4.1 | 0.9 | 0.9 |
| Total | 519 | 100.0 | 100.0 | | | |

Because of the small sample in the SOS, we are forced to return to the less sophisticated weighting method detailed in the first section of the paper. Table 12, above, compares the

demographics directly between the two surveys, and includes the preliminary and final survey weights. Application of those weights produced weighted sample demographics for the SOS that in no instance varied by more than one half of one percentage point from the SES sample demographics (table not included here). The minimum value of the weight variable was 0.48 and the maximum value was 2.46.

Once again, as shown in Table 13, below, the results show minimal changes in the survey responses. To the extent that I had a hypothesis coming in, it was that the classroom administration might have hurt our numbers overall. Table 13 shows minimal change, and what changes do occur are in the opposite direction of the one hypothesized. So once again, we are forced to accept the null hypothesis that weighting does not make a difference in the survey results.

**Table 13: UAlbany 2006 SOS: Weighted to SES Demographics**

| | | Unweighted | Weighted |
|---|---|---|---|
| # | Question/Response | n=519 | n=520 |
| Si1 | **Academic experiences have:** | | |
| | Not met expectations (1) | 13.5 | 14.0 |
| | Met expectations (2) | 73.0 | 72.7 |
| | Exceeded expectations (3) | 13.3 | 13.3 |
| | Average | 2.00 | 1.99 |
| Si3 | **Would choose UAlbany again:** | | |
| | Definitely No (1) | 5.2 | 5.1 |
| | Probably No (2) | 9.6 | 10.1 |
| | Uncertain (3) | 17.2 | 17.5 |
| | Probably Yes (4) | 39.1 | 38.8 |
| | Definitely Yes (5) | 28.8 | 28.4 |
| | Top Two Categories | 67.9 | 67.2 |
| | Average | 3.77 | 3.75 |
| Si6 | **Quality of Education is:** | | |
| | Very Low (1) | 0.6 | 0.5 |
| | Low (2) | 1.9 | 2.2 |
| | Average (3) | 42.2 | 43.3 |
| | High (4) | 49.1 | 47.8 |
| | Very High (5) | 6.2 | 6.2 |
| | Top Two Categories | 55.3 | 54.0 |
| | Average | 3.58 | 3.57 |
| Si7 | **Overall Satisfaction:** | | |
| | Very Dissatisfied (1) | 0.8 | 0.8 |
| | Dissatisfied (2) | 6.6 | 7.1 |
| | Neither Sat. nor Diss. (3) | 16.1 | 15.7 |
| | Satisfied (4) | 61.3 | 61.1 |
| | Very Satisfied (5) | 15.3 | 15.3 |
| | Top Two Categories | 76.6 | 76.4 |
| | Average | 3.84 | 3.83 |

The 2009 Student Opinion Survey

After having conducted these analyses, and after consulting with the other SUNY university centers, who were all conducting the survey online as well, we decided to go ahead and administer it online for the Spring, 2009 survey. The final section of this paper will thus address the question of whether the web administration resulted in a de facto sample substantially different from the overall student body to impact the survey results (and potentially the inter-SUNY rankings).

To test this possibility, I again weighted the survey data to population parameters. Table 14, below, shows that, as with the previous surveys, survey respondents differed substantially from the population with regard to race and gender. As with previous surveys, students admitted as freshmen had proportionately higher representation than students admitted as transfers, but this time there were no important differences by student level. As a result, I weighted for race and gender and admit type, but not student level. Table 14 shows that after weighting, the sample is essentially representative of the population on all parameters shown.

Tables 15 and 16, below, show that, once again, weighting did not change the results of the survey, either from the perspective of substantive importance or statistical significance. Table 15 shows the results of a number of questions that get at general satisfaction, and in no instance did weighting change anything at any material level; in fact, on three out of four items shown, the percentages would have been identical had I rounded to the nearest full percentage point, as is the norm with these types of survey results.

Table 16 shows the same pattern with a set of topical questions. For three out of five questions weighting would have resulted in no difference after rounding, and in no instance was the difference larger than that. Interestingly, looking at the averages (the items used by SUNY for comparative purposes) four out of the nine items have the same averages; four have slightly higher averages when weighted, and one has a slightly lower average when weighted. Since SUNY uses unweighted figures, we can feel confident based on these analyses that we were not artificially improving our numbers by shifting to a web-based survey. If anything, we might be slightly better off with the old classroom sample (although again, the differences are truly small).

**Table 14: UAlbany 2009 SOS: Weighted Demographics**

| Female | *Unweighted* Frequency | Sample Percent | *Weighted* Frequency | Sample Percent | Population Percent | Difference |
|---|---|---|---|---|---|---|
| White | 687 | 35.3% | 515 | 26.5% | 26.5% | 0.0% |
| Black | 114 | 5.9% | 109 | 5.6% | 5.6% | 0.0% |
| Hispanic | 99 | 5.1% | 82 | 4.2% | 4.3% | -0.1% |
| Asian or Pacific Islander | 65 | 3.3% | 57 | 2.9% | 2.9% | 0.0% |
| Amer. Indian or Alaska Nat. | 1 | 0.1% | 1 | 0.1% | 0.1% | 0.0% |
| Non-Resident | 25 | 1.3% | 24 | 1.2% | 1.2% | 0.0% |
| Unknown | 204 | 10.5% | 150 | 7.7% | 7.8% | -0.1% |
| Total, Female | 1195 | 61.4% | 938 | 48.2% | 48.4% | -0.2% |
| **Male** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| White | 467 | 24.0% | 595 | 30.6% | 30.4% | 0.2% |
| Black | 40 | 2.1% | 77 | 4.0% | 3.9% | 0.1% |
| Hispanic | 51 | 2.6% | 74 | 3.8% | 3.8% | 0.0% |
| Asian or Pacific Islander | 52 | 2.7% | 57 | 2.9% | 2.9% | 0.0% |
| Amer. Indian or Alaska Nat. | 0 | 0.0% | 0 | 0.0% | 0.1% | -0.1% |
| Non-Resident | 12 | 0.6% | 28 | 1.4% | 1.3% | 0.1% |
| Unknown | 128 | 6.6% | 177 | 9.1% | 9.0% | 0.1% |
| Total, Male | 750 | 38.5% | 1008 | 51.8% | 51.5% | 0.3% |
| Total, Sample | 1946 | 100.0% | 1947 | 100.0% | 100.0% | 0.0% |
| **Student Level** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 292 | 15.0 | 282 | 14.5% | 15.3% | -0.8% |
| Sophomore | 464 | 23.8 | 460 | 23.6% | 23.7% | -0.1% |
| Junior | 596 | 30.5 | 597 | 30.6% | 29.8% | 0.8% |
| Senior | 600 | 30.7 | 611 | 31.3% | 31.1% | 0.2% |
| Total | 1952 | 100.0 | 1953 | 100.0% | 100.0% | 0.0% |
| **Admission Type** | ***Unweighted* Frequency** | **Sample Percent** | ***Weighted* Frequency** | **Sample Percent** | **Population Percent** | **Difference** |
| Freshman | 1337 | 68.5 | 1254 | 64.2% | 65.1% | -0.9% |
| Transfer | 615 | 31.5 | 699 | 35.8% | 34.9% | 0.9% |
| Total | 1952 | 100.0 | 1953 | 100.0% | 100.0% | |
| **Additional Demographics** | ***Unweighted* Percent** | | ***Weighted* Percent** | | **Population Percent** | **Difference** |
| Age | 22 | | 22.0 | | 22.0 | |
| UAlbany GPA | 2.93 | | 2.9 | | 2.8 | |
| Transfer Admits | 31.5% | | 35.8% | | 34.9% | |
| Full-Time | 96.8% | | 96.8% | | 94.2% | |
| On-Campus Residence | 62.2% | | 60.5% | | 56.3% | |
| Weight Variable Minimum: | 0.71 | | | | | |
| Weight Variable Maximum: | 2.40 | | | | | |

**Table 15: SOS 2009 General Satisfaction, Weighted vs. Unweighted**

| Survey Question | Unweighted | Weighted | Difference |
|---|---|---|---|
| **UAlbany Met or Exceeded Academic** | 83.2% | 82.9% | 0.3% |
| Average | 2.01 | 2.01 | 0.00 |
| **UAlbany was 1st or 2nd Choice** | 76.7% | 78.2% | 1.5% |
| Average | 1.87 | 1.84 | 0.03 |
| **Prob. or def. would Choose UAlbany again** | 67.9% | 68.2% | -0.3% |
| Average | 3.80 | 3.81 | -0.01 |
| **Satisfied or Very Satisfied with UAlbany in** | 74.8% | 75.0% | -0.2% |
| Average | 3.80 | 3.80 | 0.00 |

**Table 16: SOS 2009 Topical Areas, Weighted vs. Unweighted**

| Survey Question | Unweighted | Weighted | Difference |
|---|---|---|---|
| **Frequently had discussions w/ instructors outside of class** | 27.0% | 27.2% | -0.2% |
| Average | 2.95 | 2.96 | -0.01 |
| **Frequently collaborated w/ other students** | 40.8% | 40.5% | 0.3% |
| Average | 3.24 | 3.24 | 0.00 |
| **Satisfied, Personal Safety/ Security on Campus** | 57.8% | 59.3% | -1.5% |
| Average | 3.47 | 3.51 | -0.04 |
| **Satisfied, Freedom from Harrassment** | 79.7% | 80.6% | -0.9% |
| Average | 4.06 | 4.08 | -0.02 |
| **Satisfied, Racial Harmony on Campus** | 71.1% | 71.4% | -0.3% |
| Average | 3.89 | 3.89 | 0.00 |

Conclusion: Why Weight?

A few important caveats bear mentioning here. First of all, I did not weight by every factor for which I could have weighted. Other factors might exist that would have produced different results. In addition, I only showed a small and no particularly random selection of items from the surveys; it is possible that other items might show more change due to weighting than did the ones I selected. On top of that, I did not do a perfect job of weighting; it is possible, if unlikely, that a more expert weighting job might have produced results more divergent from the original unweighted survey samples.

Finally, while weighting may indeed correct for non-representativeness of the survey "sample," it is impossible to correct for non-response bias unrelated to the factors

included in the weights – particularly the possibility that respondents, regardless of their characteristics, may be more engaged and have higher satisfaction levels than non-respondents. We also need to be careful not to make any particular under-represented group try to speak for a much larger group of non-respondents.

Given that all this weighting produced nothing but null findings, one might well ask: "why bother?" From a practical perspective, there might not be much apparent benefit to weighting. The results seem unlikely to change a great deal; we don't always have a lot of spare time to tinker with weights; and finally (as mentioned earlier) weighting might appear to some less informed observers like tampering with the data.

Despite all that, there are good reasons to weight university survey data:

1) Stratified Samples. The 2006 SOS sample design was essentially a stratified sample, in which the administrators got as close as they could to producing a probability sample. But the classes chosen were not a true probability sample; as we have seen, some groups had greater and some had lesser probabilities of selection. For this type of sampling methodology, weighting for probability of selection is indeed required in order to conduct any statistical tests on the data. In this case, I in effect combined weighting by differential probability of selection with weighting by nonresponse (see Groves et al., 2004, pp. 323-326).

2) You Never Know. Just because one survey didn't change after weighting doesn't mean that the next one will not. Thus, it is always worth the small amount of time it takes to try at least a quick first-stage weighting scheme to see if anything jumps out at you. If it does, you can put in the additional time and effort to really do it right; if there isn't anything there, you can tell people that you checked.

3) Not All Items are the Same. Just because some survey items don't change after weighting doesn't mean you can be sure that none will. For example, some survey items may be particularly sensitive to student level; others might be more sensitive to race; still others might be more sensitive to gender. We should always keep that in mind when thinking about weighting, and make sure we include relevant weighting variables whenever possible.

4) Campus Politics. Suppose that your campus has an undergraduate population that is 10% African American and you issue a survey report showing that in your survey, only 5% of your sample is African American. Some people might not be happy with that, and they would have a point!

5) Do the Right Thing. Finally, even if none of these other factors applied, we should still consider weighting whenever we have time to do so. We have the sample demographics; we have the population demographics; if they differ systematically, weighting is simply the right thing to do. We don't necessarily have to report the weighted results (especially when they show only minor differences), but even in these cases, weighting is still valuable insofar as it increases our confidence in the validity and reliability of our survey results.

## REFERENCES


Bloom, Joel (2008).Worth the Weight?: The Benefits and Pitfalls of Applying Post-Stratification Weights to Web Surveys of College Undergraduates. *Presented at the 2008 Annual Meeting of the American Association of Public Opinion Research, New Orleans, Louisiana.*

Couper, Mick (2008).*Designing Effective Web Surveys.* Cambridge: Cambridge University Press.

Crawford, S.D., McCagbe, S.E., Saltz, B., Boyd, C.J., Freisthler, B., and Paschall, M.J. (2004).“Gaining Respondent Cooperation in College Web-Based Alcohol Surveys: Findings from Experiments at Two Universities.” Paper presented at the 2004 annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona.

Dillman, Don A., Smyth, Jolene D., and Christian, Leah Melani (2009).*Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method.* New York: Wiley.

Groves, Robert M., Floyd J Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau (2004).*Survey Methodology.* New York: Wiley.

Heerwegh, D. (2005).Effects of Personal Salutations in E-Mail Invitations to Participate in a Web Survey.*Public Opinion Quarterly, 69:588-598.*

Joinson, A.N. (2005).“Audience Power, Personalized Salutation and Responses to Web Surveys.”Paper presented at the ESF Workshop on Internet Survey Methodology, Dubrovnik, Croatia.

Joinson, A.N., and Reips, U.-D.(2007).Personalized Salutation, Power of Sender and Response Rates to Web-Based Surveys.*Computers in Human Behavior, 23*:1372-1383.

Joinson, A.N., Woodley, A., and Reips, U.-D.(2007). Personalization, Authentication and Self-Disclosure in Self-Administered Internet Surveys.*Computers in Human Behavior, 23*:275-285.

Kaplowitz, Md., Hadlock, T.D., and Levine, R. (2004).A Comparison of Web and Mail Survey Response Rates.*Public Opinion Quarterly, 68*:94-101.

Mandell, Lewis. (1974). When to Weight: Determining Nonresponse Bias in Survey Data.*Public Opinion Quarterly, 38*:247-252.

Pearson, J., and Levine, R.A. (2003).Salutations and Response Rates to Online Surveys. Paper presented at the Association for Survey Computing Fourth International Conference on the Impact of Technology on the Survey Process, Warwick, England.

Porter, S.R., and Whitcomb, M.E. (2003).The Impact of Lottery Incentives on Student Survey Response Rates.*Research in Higher Education, 44*:389-407.

Singer, Eleanor, ed. (2006). *Special Issue: Nonresponse Bias in Household Surveys. Public Opinion Quarterly* 70:5.

Trouteaud, A.R. (2004). How you ask counts: A test of internet-related components of response rates to a web-based survey. *Social Science Computer Review, 22*:385-392.

# MEASURING SUCCESS FOR ALL STUDENTS

David A. Hemenway
Director of Institutional Effectiveness
Charter Oak State College

Sarah K. Hemenway
Graduate Student
University of California, Santa Cruz

## Abstract

We conducted a brief environmental scan of undergraduate education in the United States and determined that, while the higher education environment has changed significantly in the last forty years, the metrics for reporting students' success have not. We proposed that institutions expand the firmly established metrics of first- to second-year retention and six-year graduation rates to all degree students and report them separately for native and transfer students by full-time and part-time enrollment status. A review of literature suggested that the standard metrics that have been used to measure traditional student success work as well measuring non-traditional student success. Using these well-understood metrics for all degree-seeking students allows institutions to provide accurate and reliable measures of student success for all students.

## The Status of Higher Education

**Higher education today.** President Obama has established a national goal of leading the world with the highest proportion of college graduates by 2020. The United States has made significant gains in the percentage of adults who have earned a bachelors degree in the last 40 years. It has also made significant gains in the number of under-represented populations both enrolling in college and earning college degrees. The total fall enrollment increased 74% from 1976 to 2008. The enrollment of White students increased by 33%, while the enrollment of under-represented students increased 276%. In raw numbers, the number of White students increased 3.0 million students while the number of under-represented students increased more than 4.6 million. The largest under-represented student increases were in Hispanic and Asian students (Snyder and Dillow, 2009). The percentage of White adults over 25 years old who have earned a bachelors degree has increased from 11.6% in 1970 to 32.9% in 2009- a 21.3 percentage point increase. The percentage of Black adults who have earned a bachelors degree has increased from 6.1% to 19.4% in this same period- a 13.3 percentage point increase. The percentage of Hispanic adults who have earned a bachelors degree has increased from 7.6% in 1980 to 13.2% in 2009- a 5.6 percentage point increase (Snyder and Dillow, 2009). These statistics demonstrate that the United States' success in increasing the percentage of adults who have earned a bachelors degree has been much more significant with White students than with under-represented students, although more under-represented students than ever are enrolling in college. If the United States is to significantly increase its percentage of adults who have earned a bachelors degree, it must

improve its success with all racial groups since the non-White United States population is increasing at a much higher rate than the White population.

In the last eight years, according to the Organization for Economic Cooperation and Development, the United States has relinquished its long held position as the country with the highest percentage of adults who have earned a bachelors degree. In 1999, the United States was first in this measure, but it slid to second behind Norway in 2007. However, the story is much more grave than just losing the number one spot in this metric. Between 1999 and 2007 the United States increased its percentage of adults with a bachelors degree from 27.5% to 30.9%- an increase of 3.4 percentage points. This is the lowest percentage point increase of any of the countries in the top ten using this metric (Organisation for Economic Co-operation and Development, 2009). This raises another important question. How many students are in the pipeline to earn a college degree? Students from North America and Western Europe (including the United States) comprised roughly half of the higher education students in 1970. In 2007, they made up less than a quarter of higher education students. While the number of United States higher education students has increased significantly in the last forty years, other countries' participation rates and raw number numbers increased much more. This is particularly true in East Asia and the Pacific. In short, North America and Western Europe have expanded their higher education enrollments at a much lower rate than other areas of the world (Organisation for Economic Co-operation and Development, 2009).

Increasing the number of college graduates is a particularly important goal because numerous studies have shown that higher education achievement is necessary for today's more complex jobs. Georgetown University's Center on Education and the Workforce

produced a report forecasting that by 2018, 63 percent of jobs will require at least some postsecondary education. The report also shows that, without a significant change in course, the labor market will be short three million educated workers over the next eight years (Carnevale, Smith & Strohl, 2010).

The United States is addressing the goal of more higher education for more people in a number of ways. For example, at a recent White House event, community college staff and students discussed how to produce five million more graduates from two-year institutions over the next decade. The Lumina Foundation's Achieving the Dream initiative is an excellent example of a successful effort to improve student success in community colleges. A number of initiatives have focused on improving the success of traditional first-time full-time students. In addition, there have been a number of programs that have concentrated on improving degree attainment for adult non-traditional students, such as Lumina Foundation's goal of having sixty percent of the United States adults have higher education credentials by 2025 (Lumina Foundation, 2009). They are putting significant resources into achieving this goal, such as recently awarding 14.8 million dollars for 19 projects to improve the success rates of adult students who have completed a substantial number of college courses, but have not earned a degree (Shapiro, 2010). All of these efforts hold promise for increasing the number of college graduates in the Unites States; however, it is interesting and important to note that many of them are concentrating on non-traditional students.

**Changes in United States higher education over 40 years.** The number of undergraduate students has increased from 7,369,000 in 1970 to 16,366,000 in 2008- an increase of 122%. The number of full-time students has increased 94% while the number

of part-time students has increased 193%. In 1970, 72% of all undergraduate students were full-time. By 2008, this percentage had declined to 63% (Aud et al., 2010).

While the number of first-time full-time undergraduate students has increased 53% from 1970 to 2008, the number of other undergraduate students has increased 141%. In 1970, 22% of all undergraduate students were first-time full-time students. By 2008, this percentage had decreased to 15%. The number of United States undergraduate students has increased significantly faster than the number of traditional first-time, full-time students over the last 40 years (Aud et al., 2010).

It is interesting to note that part-time first-time freshmen enrollment has only increased 25% since 1970. The proportion of full-time to part-time freshmen enrollment has remained relatively stable since 1970 (Aud et al., 2010). This suggests that the increase in United States undergraduate enrollment is not because of an increase in traditional first time students.

In the last forty years, the undergraduate enrollment at four-year institutions has increased 90% while it has increased 211% at two year institutions. The majority of the growth in undergraduate education has been at the two-year institutions. In 1970, 69% of the undergraduates attended a four-year institution. By 2010, this percentage had declined to 57%. The number of first-time freshmen has increased more at four-year institutions than at two-year institutions since 1970. While the number of full-time students has increased 89% at four-year institutions since 1970, the number of full-time students at two-year institutions has increased 139%. The percentage of full-time undergraduates attending four-year institutions has declined from 77% in 1970 to 72% in 2010. While the number of part-time students has increased 96% between 1970 and 2010 at four-year

institutions, it has increased 291% for two-year institutions. In 1970, 48% of all part-time students attended four-year institutions. By 2010, this percentage had declined to 31%. The percentage of part-time students attending two-year institutions swelled from 52% in 1970 to 69% in 2010 (Aud et al., 2010).

**Summary.** In summary and as indicated by the tables in the "Extended Data Tables" section, the following changes have occurred in the United States higher education environment in the last forty years:

- The number of undergraduate students has increased from 7,369,000 in 1970 to 16,366,000 in 2008- an increase of 122%.
- The part-time undergraduate enrollment has increased 193% while the full-time undergraduate enrollment has increased only 94%.
- While first-time, full-time undergraduate enrollment has increased 53%, other undergraduate enrollment has increased 141%- nearly three times the rate of first-time, full-time undergraduate enrollment. First-time, full-time enrollment (while still a significant portion of undergraduate higher education) is becoming a much smaller percentage of total undergraduate enrollment.
- First-time undergraduate enrollment has increased only 47% since 1970, while total undergraduate enrollment has increased 122%.
- US higher education has become more diverse in the last 40 years.
- Two-year colleges have served the majority of the under-represented students.
- Two-year institution undergraduate enrollments have increased 211% compared to four year institutions' 90% increase. Two-year institutions enrollment have grown at more than twice the rate of four year institutions.

- First-time freshmen enrollment has increased 55% at four-year institutions as compared to 37% for two-year institutions. This indicates that four-year institutions are continuing to emphasize serving traditional students.
- Full-time undergraduate enrollment has increased 139% at two-year colleges as compared to 89% at four-year colleges. Two-year colleges have significantly expanded their full-time enrollment more than four-year colleges in the last forty years.
- Part-time undergraduate enrollment has increased 291% at two-year colleges compared to a 96% increase at four-year institutions. Part time enrollment has increased at two-year colleges three times faster than it has at four-year institutions.

## Retention Measures: Past and Present

**Efforts to measure student success.** The first reference we could find for a student success metric used by the Federal government was in the 1965 Higher Education Act, which established the Higher Education General Information Survey (HEGIS) that included a graduation rate measure. We believe that graduation rate measures were first reported in the late 1960s.

The two measures of student success were defined as the first- to second-year retention rate and the graduation rate, which was calculated by the percentage of students who completed their degrees within 150% of the normal time to complete a degree (defined as three years for an associates degree and six years for a bachelors degree). These metrics were used to measure the success of first-time, full-time degree seeking

students and, at the time they were developed, these students represented 22% of all undergraduate students (Aud et al., 2010).

Indicators of student success came to the forefront as a result of the Student Right-to-Know Act, which became law in 1990. This law required institutions to report the completion or graduation rates of certificate or degree seeking full-time students for each academic program. This law also required the reporting of the success of student athletes. The graduation rates part of IPEDS was developed specifically to help institutions respond to these requirements ("Student Right-to-Know Act," n.d.). The graduation rate was determined by the Student Right to Know Act as the total number of completers who completed their academic programs within 150% of the normal time required to complete the program by the revised adjusted cohort which was defined as the number of students who entered the institution minus some students who could be removed because of death, or other established reasons ("Completer," n.d.). While this law was enforced rigorously in some parts of the country, it was not enforced consistently across the country.

One of the first successful efforts to measure student success at different institutions was the Consortium for Student retention Data Exchange (CSRDE) which was established at the University of Oklahoma in 1994. This project includes both two-year and four-year institutions and their goal is to provide "timely, comprehensive, comparable benchmarking data" on student success. This project only includes traditional first-time full-time degree seeking students (CSRDE, 2010).

CSRDE presents retention/graduation information for cohorts of students that are measured yearly for 8 to 10 years. Member institutions can get peer institution information for benchmarking their student success measures with other institutions. The

advantage of the CSRDE information is that it is comparable between institutions because their measures of student success include comparable students (i.e., traditional full-time, first-time degree seeking students). This is a very valuable resource for benchmarking student success. However, its weakness is that it doesn't provide any student success information for non-traditional students, which make up the majority of undergraduate students in the United States.

CSRDE presents information on a cohort of first-time, full-time degree seeking students for up to ten years after they entered the institution. This allows us to understand when these students are retained or graduated. For one large state university the 6-year graduation rate was 40.5%, the 7-year rate was 43.5%, the 8 year rate was 44.9% and the 9 year rate was 46.1%. The difference between the cumulative 6-year rate and the 9-year rate is only 5.6 percentage points. This seems to be fairly typical for many institutions: The success rates for first-time full-time degree seeking undergraduate students is best measured at the six year time because by then the majority of the students who are going to graduate have graduated from the institution. While the six-year measure does not capture the success of all students, it does reasonably reflect the success rate of these students.

In recent years, a number of organizations have included student success measures in their materials. The Education Trust is one such organization. The goal of the Education Trust is to "promote high academic achievement for all students at all levels- pre-kindergarten through college. Our goal is to close the gaps in opportunity and achievement that consign far too many young people- especially those from low income families or who are black, Latino, or American Indian- to lives on the margins of the

American mainstream" (The Education Trust, 2010). The Education Trust, which has received significant funding from the Lumina Foundation, has developed reporting tools, which presents metrics of student success. These include the first to second year retention rate, and four year, five year and six year graduation rates. These indicators of student success are presented for different races/ethnicities for a number of years. While these rates are calculated for only first-time, full-time students, they are presented as if they are appropriate to all students. Consumers are misled by this information because they are comparing the success rates of different institutions without being told the constraints and limitations of these student success measures.

The National Association of System Heads (NASH) and the Educational Trust worked together to establish the Access to Success Initiative (A2S). Twenty-four state higher education systems enrolling more than three million undergraduate students were part of this effort. One of the goals of this effort is to measure student success by income and underrepresented minority status (URM) for traditional and non-traditional students. This group developed a number of metrics including first to second year and six-year graduation rates for traditional and non-traditional students. Traditional students were first-time, full-time, degree seeking students and non-traditional students were transfer students who were first-time degree seeking students at their bachelors college. This project directly compared transfer (i.e. non-traditional) students with freshmen students (i.e. traditional) students (The Education Trust, 2010). The Connecticut State University System was part of this project. The metrics they used allowed a direct comparison between transfer and native students and they found that transfer students were more successful than native students in first to second year retention and in six-year graduation

rates. This project was one of the first to include student success metrics for transfer students. This was a very worthwhile addition to their reporting of student success. However, this project did not consider the enrollment status (full-time versus part-time) of the non-traditional students. This addition would have allowed a more complete analysis of the project's data since full-time transfer students have different enrollment patterns than part-time transfer students (National Association of System Heads, 2009).

The College Portrait project of the Voluntary System of Accountability (VSA) has metrics for student success for traditional and non-traditional students. The Voluntary System of Accountability (VSA) Online is an initiative by public 4-year universities to supply basic, comparable information on the undergraduate student experience to important constituencies through a common web report: the College Portrait. The VSA was developed in 2007 by a committed group of university leaders and is sponsored by two higher education associations- the Association of Public and Land-grant Universities (APLU) and the Association of State Colleges and Universities (AASCU). The development and start-up funding was provided by the Lumina Foundation (VSA, 2010a).

The College Portrait presents first to second year retention rates for traditional students (first-time, full-time, degree seeking students) and for transfer students (first time at the transfer college, full-time, degree seeking). One of the interesting facts presented by the College Portrait is that the first to second year metric for traditional full-time students are comparable to those for non-traditional or transfer students. This is the first time the authors have seen a direct comparison between the success of traditional native students and transfer students.

In addition, the VSA project provides the six-year graduation rates for traditional and transfer full-time degree seeking students. Consumers can compare the success rates of traditional and transfer full-time students at different colleges. One weakness of the VSA is that it does not provide the metrics for part-time non-traditional students. This is unfortunate because many degree-seeking students enroll on a part-time basis (VSA, 2010).

Transparency by Design began as a project of the Presidents Forum, whose mission is to advance the recognition of innovative practice and excellence in serving adult students. This organization established the College Choices for Adults project in 2008, which is funded by the Lumina Foundation. It currently includes seventeen online colleges serving adult students. This organization is comprised of institutions in the non-profit and for profit sectors of higher education. This project currently provides institutional demographics, academic program information, student engagement information from the National Survey on Student Engagement (NSSE) and evaluation information from graduates. This group recently developed their metrics for measuring student success. They will be using a first to second year metric for measuring learner retention and another metric for measuring learner completion, which will be measured at 150% and 200% of the normal time to complete a degree. All degree-seeking students will be included in the cohorts. The cohort information will not be presented by full-time and part-time student status. This was discussed while these metrics were being developed and the group decided that student enrollment status was not necessary. They did decide that transfer would be considered a positive outcome and would be included in the analysis of the cohort. These metrics of student success will be implemented in the next

year and the results will be presented on their website, College Choices for Adults (Transparency by Design, 2010; D. Hemenway, personal communication).

In June 2010, the National Governors Association (NGA) has produced a report entitled ”Complete to Compete, Common College Completion Metrics.” This report is the product of a workgroup that the NGA established to make recommendations on the common higher education measures that states should collect and report publicly. The NGA felt that common college completion metrics are “essential” for states under the current fiscal constraints. Politicians want to be able to compare higher education efficiency with those of other state provided services using cost/benefit ratios and production ratios. They feel this is necessary to ensure that their investments in higher education are producing reasonable returns on their investments. The NGA is recommending that states develop appropriate unit record student tracking systems that have a unique statewide student identifier, have student level data for all public colleges and universities on enrollment, demographics, financial aid, transfer, persistence, course/transcript information including enrollment in developmental/remedial courses, and degree completion information. They want to ensure that there is privacy protection for all individually identifiable student records and there is a data audit system to ensure data quality, validity, and reliability (Reyna, 2010).

Achieving the Dream is a large Lumina Foundation project whose primary goal is to help more community college students succeed (earn degrees, earn certificates, or transfer to other institutions to continue their studies). One of its main strategies to achieve this broad student success goal is to encourage community colleges to build a culture of evidence that can be used to make decisions on how the colleges can improve

their services to students at risk. The project started in 2003 and has expanded rapidly to 80 institutions in 15 states (Jenkins, Ellwein, Wachen, Kerrigan, & Cho, 2009).

Student success measures were developed beyond the traditional first-time, full-time cohort. Achieving the Dream measures the success of all first-time degree- and certificate-seeking students and presents its results by students' full-time and part-time enrollment status. One of the Achieving the Dream policy briefs compared community college student achievement in six states (Connecticut, Florida, North Carolina, Ohio, Texas, and Virginia). In this study, student success was determined by the number of students who earned a degree, the number who transferred without earning a degree and the number who were still enrolled at their community college after earning thirty or more credits. With this expanded definition of success, the study showed the percentage of students who succeeded at the community colleges in these six states ranged from 33% to 51%. These percentages are significantly higher than the traditional success measures calculated for first-time, full-time degree-seeking students. This study also showed that part-time degree- and certificate-seeking students achieved their goals at a significantly lower level (more than 15 percentage points lower) than full-time degree- and certificate-seeking students in all six states (Achieving the Dream, 2008).

There has been much interest in measuring student success in the last twenty years. The efforts started with the Student Right to Know Act of 1990 and have continued thru the efforts of the Consortium for Student Retention Data Exchange (CSRDE), the Education Trust's College Results Online project, the National Association of System Heads (NASH) Access to Success Initiative (A2S), the Association of Public and Land-grant Universities (APLU) and the Association of State Colleges and Universities

(AASCU)'s Voluntary System of Accountability-College Portrait, Transparency by Design's College Choices for Adults, the National Governor's Association (NGA) Common College Completion Metrics project, and the Lumina Foundation's Achieving the Dream project. What have we learned from all these projects? We have learned the following about student success measures:

**Justification for current retention measures.** First, before we discuss modifications of these measures to include non-traditional students, let us discuss why we use these measures. First to second year retention rate is a frequently used retention measure, as approximately 75% of departed students leave during the first year (Tinto, 1987) and, each year, fewer students are returning for their second year at traditional four-year institutions (ACT, 2009). It has also been noted that the reasons that students leave during the first year tend to be completely separate from the reasons that students leave at any other point in their college education (Tinto, 1988). This particular section will focus on the psychosocial factors that lead to first year student departure before their second year, particularly expectations of the college experience, emotional support and relationship with previous community, identity transformation, and self-efficacy and performance goals. For an excellent review of the overall factors that relate to student retention, please see Campbell and Mislevy (2009).

For the first year student moving from high school (or the career world) to college, there is frequently a mismatch between the student's expectations of college life and the reality of college life. High schools and colleges do very little collaboration to ease the transition, instead opting to leave students to their own devices (Kirst & Venezia, 2004). Stanford University's Bridge Project began in 1996 and investigated the transition

between high school and college in six states (California, Illinois, Georgia, Maryland, Oregon, and Texas) through interviews and surveys administered to administrators, students, and parents at high school, community college, and four-year university level. Relevant to our purpose here, high school students take multiple assessments, including the PSAT, the SAT, the AP and so on, that have varying levels of influence on their admission to college, but little consistent influence on students' placement in appropriate college coursework, which commonly leads to a repetition of past coursework or placement at too high of a level where the student becomes lost. Additionally, students proceed through their high school education by taking tests and mastering subject matter, whereas college professionals indicate a desire for students with the ability to critical think, a skill not necessary taught or evaluated in high school. This leads to students believing that meeting high school graduation requirements is adequate preparation to allow them to succeed at the college level. The study also identified that students enter college believing that they don't have to worry about their grades until their second year, they can take whatever classes they want, and that getting into college is the hardest part. These are all expectations that are quickly challenged upon the student's arrival at college (Venezia, Kirst, & Antonio, 2003). Upon comparing 31 first-year students' expectations of college with those same students' experiences at the middle and end of their first year, not only did students' expectations significantly differ from their experiences, but students with unrealistic expectations had lower first year grade point averages (Smith and Wertlieb, 2005).

As some students begin their college career, they receive an immense amount of support from their home community (teachers, family, friends, significant others),

whereas others do not. Additionally, some students feel the need to reject the belief systems of their previous community in order to become integrated in their new college community and succeed, as opposed to maintaining their previous belief system, whether or not it allows them to integrate into the new community. Both a lack of support and a lack of perceived need to reject previous attitudes and values increase the probability that a student will depart from an institution before their second year. Support and encouragement was particularly key from high school educators, parents, friends, and significant others (Nora and Rendon, 1990; York-Anderson and Bowman, 1991; Attinasi, 1989). Support was especially critical for first-generation students, as having family and friends who did not have experience with college led to a lacking support structure for the students (Hsaio, 1992). The perceived need to reject previous attitudes and values is a core step in the college student's process of separation from their previous community as they begin their college journey (Tinto, 1987). An exploration of first-to-second-year persistence by first-time, full-time freshmen at a public, four-year institution through three different data collection tools investigated the direct and indirect effects of student pre-entry characteristics, initial institutional commitment, separation, and first-to-second year persistence, with student pre-entry characteristics operationalized to support and rejection of attitudes and values. Both support and rejection of attitudes and values were significantly associated with persistence to the second year (Elkins, Braxton & James, 2000).

An "identity crisis" is a psychological phenomenon noted by psychologist Erik Erikson that is a time of intense reflection and potential change in an individual's view of him- or herself. He notes that the "identity crisis" is one of the most important issues a

young adult faces in his or her development through the teenage years, which coincides with the typical transition time between high school and college (Erikson, 1970). The move from high school to college is frequently seen as a time when a teenager reaches independence, a transition that leads to a change of identity that may involve changing existing identities, adding new ones, or leaving others behind. A qualitative research project looking into college-bound high school seniors and their parents during the college admission process revealed that students see leaving for college as a time to discover who they "really" were and that the identity they establish in college is the one that will stay with them for the rest of their life. The students involved in the study frequently used terms such as a "fresh start" when viewing their transition to college. The identity crisis that students face as they begin their college journey is just one of any new social and emotional stressors that first year students face that contributes to their decision to continue to their second year (Karp, Holmstrom & Gray, 1998; Smith & Wertlieb, 2005).

Social Cognitive Career Theory suggests that a student's beliefs about items such as self-efficacy (confidence in academic ability), expectations (the consequences of graduating from college), and performance goals (motivation to graduate from college) all relate to the student's performance of a behavior such as remaining in college (Lent, Brown, and Hackett, 1994). An analysis of the literature connecting both Social Cognitive Career Theory and task persistence is available in Kahn and Nauta (2001). A study at a large midwestern university assessed each of these three beliefs through a variety of measures and associated them with first to second year persistence. The three beliefs were assessed both before the students began college and during their second

semester. The results indicated a strong correlation in the social-cognitive factors measured during the second semester whereas there was no significant correlation in the social-cognitive factors measured prior to the students beginning their college career. While this doesn't provide a predictor of student persistence before the student begins college, it does indicate that, after some experience with college, social-cognitive factors are strong predictors on whether the student will continue to the second year (Kahn & Nauta, 2001).

Many of these factors appear to be exclusive to the first-time, full-time student who entered college directly after graduating from high school and not immediately translatable to adult and otherwise non-traditional students. In 1994, adult students constituted approximately half of the college population and present a different set of concerns and stressors that influence their persistence into their second year of education (MacKinnon-Slaney, 1994). While not specific to first to second year retention, the Adult Persistence in Learning Model (MacKinnon-Slaney, 1994) attempts to address these unique concerns. The model includes three components: personal issues, learning issues, and environmental issues. We will focus mainly on personal issues in this paper, to continue the trend of psychosocial issues that result in low first to second year retention. An additional complete review of factors involved (and not involved) in adult student persistence can be found in Bean and Metzner (1985).

The five factors of the personal issues component are: self-awareness, willingness to delay gratification, clarification of career and life goals, mastery of life transitions, and sense of interpersonal competence. At first glance, the factors of the personal issues component closely mirror similar factors for first-time full-time students; however, the

implications of these factors are somewhat different for the adult student than they are for a first-time full-time student. For instance, while persistence is increased for first-time full-time students when they are open to a shift of their identity and a rejection of previous community beliefs, increased persistence in adult students is connected with self-awareness, which the author defines as requiring a "robust sense of self." While first-time full-time students are generally facing their first major life transition as they head to college, adult students must have already mastered this skill in order to be able to transition into the higher education environment successfully. Further review of the Adult Persistence in Learning Model can be found in MacKinnon-Slaney (1994).

As indicated earlier, integration into the new university environment is important for first-time full-time students in order to be retained into their second year (Tinto, 1987). This plays a similar, but more complicated role for adult students who commonly do not reside on campus and have responsibilities outside of their education. Research into management majors at an academic center in a major metropolitan area revealed that social integration factors such as similarity to one's classmates resulted in higher persistence, although the author cautions that the particular group studied was mainly focused on career development as opposed to intellectual growth (Ashar & Skenes, 1993). The isolation that adult students face from feeling like they do not belong at the university with first-time full-time students can also lead to poor persistence (Metcalf, 1993).

Support is another area of similarity between first-time full-time students and adult students. Similar to the results for first-time full-time students, adult students who don't

have a support structure in their family have lower retention rates than those that do (Comfort et al., 2002).

A common assumption is that adult students are generally faced with more life events and more commitments outside of school than the average first-time full-time direct entry college student. On an intuitive level, this assumption makes rational sense. One hundred and twenty one first-time distance education students between the ages of 30 and 45 took a series of questionnaires such as the Resiliency Attitudes Scale, the Life Events Inventory, and a survey on external commitments. These responses were then connected to their student record data regarding the completion and non-completion of coursework, which the author used to determine persistence. External commitments (specifically, work commitments) were the only significant factor between the two groups, with those with higher work commitments tending to have lower persistence. None of the other factors reached significance. While this isn't specifically a measure of first to second year retention, the adequate completion of coursework is certainly an influential factor on a student's persistence in education.

In conclusion, the research shows that students who withdraw from college are most likely to do so between the first and second years for the reasons outlined above. The majority of the research is dedicated to traditional (e.g., first-time, full-time) students, but the body of research suggests that both traditional and non-traditional students face similar challenges when adapting to the rigors of college life. The process of learning an institution's ways and processes is very similar for both groups of students; however, non-traditional students are frequently faced with additional challenges, such as learning to balance between work, home, and school. This research supports the validity of

measuring student retention between the first and second years of enrollment because this metric provides an excellent predictive measurement of the student's commitment to their achievement of their higher education degree. In addition, it allows institutions to measure the effects of their first year student orientation processes, whether they be geared toward traditional or non-traditional students (or both!).

The idea behind measuring first to second year retention is as valid for non-traditional students as it is for traditional students. Numerous studies have shown that one of the most important factors in student success is engaging students in their college and easing the challenges of learning how to succeed at their new college. While traditional students and non-traditional students are different in many ways, they are similar in their need to acclimate to their college. First to second year metrics measure student success at this crucial time for student achievement. This has been shown in the discussion above.

**The problem with current retention measures.** Higher education still uses the same two basic measures of student success that it developed over forty years ago when the higher education environment was very different than it is today. While the higher education environment has changed to meet the needs of society, our basic measures of student success have not. We are still using the first to second year retention rate and the six-year graduation rate for first-time full-time students as our yardstick for success in higher education even though traditional first-time, full-time students have become the minority in the higher education student population.

Traditional measures of student success were developed for measuring the success of first-time, full-time, degree-seeking students when this type of students represented the vast majority of college students. The two major metrics used to measure student success

are the first to second year student retention rate and the six-year graduation rate. The first to second year retention rate is a very important measure of student success because the majority of students who fail to graduate from a college withdraw before the second year. The six-year graduation rate is an excellent indicator of student success because the majority of first-time, full-time, degree-seeking students complete their degrees within the six-year period if they earn their degree from their original college. These two metrics provide a reasonable measure of student success for traditional students and have been used to benchmark institutional achievement for many years. These measures have become commonplace in the literature and have been used by policy makers and consumers for many years to evaluate colleges.

The problem is that these metrics do not measure the success of non-traditional students. For example, students who begin their studies at community colleges and then transfer and students who drop out of college and then return to different college are by definition excluded from these measures because they were not first time, full-time, degree seeking students at the college where they earned their degrees. This has become a growing problem as more college students have taken non-traditional paths to attain their college degrees. In addition, innovative colleges have utilized non-term based instruction, which makes it impossible to utilize the term-based approach to measuring student success.

At Charter Oak State College we have been measuring student success with a first to second year metric for a number of years. Since Charter Oak State College matriculates students anytime, the term-based approach to measuring student success does not work for us. Charter Oak State College charges a matriculation fee that is renewed annually

until a student completes her/his degree. We measure the number of students who enter during any month and calculate how many are retained or graduated thirteen months later. This metric has allowed the College to measure its success with engaging students into its program. Using this model, approximately seventy to seventy-five percent of our first time, degree-seeking students are retained or graduated in the first thirteen months of enrollment and there is no significant difference between the ethnic groups of students. The majority of the students who leave Charter Oak State College do so within the first year. It turns out that the first to second year measure of student engagement works as well in a non-traditional college as it does for a traditional college if we broaden the definition of the metric to accommodate the way a non-traditional college operates.

One of the concepts behind the traditional six-year graduation rate is that nearly all first-time, full-time, degree seeking students graduate within six years after their matriculation into their college. We have found this to be true for Charter Oak State College! Charter Oak State College's metric for degree completion is six years for our students pursuing a bachelors degree. Approximately fifty percent of our first-time, degree-seeking students complete their bachelors degree within six years. There is no significant difference between ethnic groups on their six-year degree completion rates. In addition, nearly all (over eighty percent) matriculated Charter Oak State College students who complete their bachelors degrees do so within six years. So, once again, we find that the "traditional" metric for measuring student success works in a non-traditional college if the definition can be adapted to meet the college's way of educating students. This information has been used internally to measure Charter Oak State College's student success and it has been used externally in the Connecticut Department of Higher

Education's Annual Accountability Plan. It has successfully served both purposes for a number of years.

**Summary**. What have we learned from our review of methods of measuring student success?

- Measuring student success is very important to many people including students, parents, college administrators, policy makers and politicians.
- First to second year retention/graduation rates and six-year graduation rates are the most commonly accepted measures for bachelors level institutions. They are familiar and understood by all constituencies.
- First to second year retention rates are important because the majority of students who withdraw from a college do so between the first and second years. A review of the literature shows there are psychological reasons for this. This measurement is as valid for non-traditional students as traditional students.
- Measuring graduation rates at the six-year level for bachelors degree institutions work because most graduates complete their degrees within that time period. This is true for traditional and non-traditional students.
- Full-time students and part-time students achieve their degrees at different rates.
- Current metrics based upon first-time, full-time degree seeking students have always been problematic and are getting more so as the undergraduate higher education environment has evolved.
- Generalizing measures based upon first-time, full-time degree seeking students to all students is commonplace, but methodologically inappropriate.

- Currently, there is no one established series of metrics that measure student success for all students. Prospective students do not have reasonable metrics to evaluate their college choices and higher education policy makers do not have good degree success measures for all types of higher education institutions.
- Many student success metrics are overly complex and not easily understood.
- In the current environment, increased scrutiny of costs and productivity in higher education will continue. If higher education doesn't develop commonly accepted performance measures, others will develop them and impose them on the higher education community.
- Success measures can and should be developed for non-traditional students.
- Metrics that work internally and externally have the greatest value.
- Student success metrics need to be broadly defined to allow the reasonable measurement of first to second year retention/graduation and six-year bachelors degree completion.
- Metrics must be statistically and methodologically sound and produce valid and reliable results for all students.

## An Expanded Model of Measuring Student Success for All Students

We propose that the higher education community build upon the current model of measuring student success. First- to second-year and six-year graduation rate measures have proven to be very effective in measuring the success of students. These metrics make sense and are understandable to the consumers of this information. While they have been applied to only first-time, full-time degree-seeking students, we propose that they be extended to all students.

This information should be calculated for all students by enrollment status (full-time and part-time) for all degree-seeking students. This would build upon the successful metrics that have been used for measuring student success for traditional students and provide reasonable measures for non-traditional students. These metrics are easily calculated for different demographic segments of the undergraduate students. While the definitions will have to take into consideration the different types of instructional approaches, we believe that reasonably comparable metrics can be produced for most institutions. We have presented sample data tables for these revised metrics in the "Sample Data Tables" section.

For example, the first to second year retention/graduation metric can be flexible enough for term-based colleges and continuous enrollment institutions. Each college will have to define its metric for first to second year retention/graduation. While these metrics maybe slightly different to meet the needs of their institutions, the basic statistic should be comparable among all institutions because they are measuring roughly the same thing. Term based colleges could use terms to measure their students' success while continuous enrollment colleges could measure their first to second year student metrics by dividing by the number of students who are retained/graduated one year later by the number of entering degree seeking students from the previous year. The most important concern is that a college establishes its methodology for measuring first to second year retention/graduation and continues to use that methodology so its information is consistent and reliable.

Data should be presented by enrollment status since full-time students and part-time students have different enrollment patterns and success rates. Comparing metrics between

different colleges is problematic if these data are not presented by student enrollment status because these statistics can mislead prospective students. For example, if a prospective student is comparing institution A with institution B and the first institution has a first to second year retention rate of 75% and the second institution has a 50% rate can they assume that the first institution is a better institution? No! The second institution may be primarily a part-time serving institution while the first maybe primarily a full-time student serving institution. Since part-time students have a different retention rate than full-time students, prospective students can be misled by incomplete information. Producing first to second year retention rates by student enrollment status can reduce this problem. When this is done, prospective students can compare apples to apples in making their college choice decisions. In addition, this metric is much more useful and valid to institutional decision makers who are using it to evaluate their colleges.

The six-year graduation rate metric can be defined as the number of students who graduated within six years of entry divided by the number of degree seeking students who entered in that given year. This approach has been successfully used in varying forms by both the VSA and the A2S initiatives. This metric can be improved by providing information for full-time and part-time students. These metrics would be useful for colleges examining their student success rates and provide valuable consumer information for individuals considering enrolling in our colleges.

This approach would be a very cost-effective solution to producing meaningful metrics for all students because our colleges are accustomed to providing this information for their traditional students. To produce this information would not require significant

new investments in our reporting systems. Our student information systems have the necessary data in their databases. No new costly data systems are required to produce this important information. The only cost is the allotment of the necessary staff time to analyze the data to turn it into useful information. These costs are minimal because we have already collected the requisite data and have developed most of the necessary programming.

## Conclusion

In conclusion, we should expand the proven metrics colleges have used to measure traditional students' success by applying student success metrics to non-traditional students. If we analyze this information by entry status (native and transfer) and enrollment status (full-time and part-time), we can produce meaningful and reasonably comparable information that will help us manage our institutions and inform the population we serve with better student success information. While these metrics would not be perfect, they would allow us to measure the effectiveness of our colleges and for prospective students who would benefit from better retention/graduation measures. These expanded metrics would allow us to measure the success of nearly all of our undergraduate students.

**Sample Data Tables**

*Table 1. First to Second Year Retention/Graduation Rate for First-Time Degree Seeking Students*

| | Reporting Year | | | | |
|---|---|---|---|---|---|
| | 2006 | 2007 | 2008 | 2009 | 2010 |
| Part-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |
| Full-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |
| All Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |

*Table 2. Six Year Graduation Rate for First-Time Degree Seeking Students*

| | Reporting Year | | | | |
|---|---|---|---|---|---|
| | 2006 | 2007 | 2008 | 2009 | 2010 |
| Part-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |
| Full-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |
| All Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |

*Table 3. First Year to Second Year Retention/Graduation Rate for Transfer Degree Seeking Students*

| | Reporting Year | | | | |
|---|---|---|---|---|---|
| | 2006 | 2007 | 2008 | 2009 | 2010 |
| Part-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |
| Full-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |
| All Students | | | | | |
| Number in Cohort | | | | | |
| Percent Retained/Graduated | | | | | |

*Table 4. Six Year Graduation Rate for Transfer Degree Seeking Students*

| | Reporting Year | | | | |
|---|---|---|---|---|---|
| | 2006 | 2007 | 2008 | 2009 | 2010 |
| Part-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |
| Full-time Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |
| All Students | | | | | |
| Number in Cohort | | | | | |
| Percent Graduated | | | | | |

## Extended Data Tables

Table 1. *Percentage of adults with bachelor's degrees by race*

| Year | White | Black | Hispanic | Total |
|---|---|---|---|---|
| 1970 | 11.6% | 6.1% | dna | 11.0% |
| 1980 | 18.4% | 7.9% | 7.6% | 17.0% |
| 1990 | 23.1% | 11.3% | 9.2% | 21.3% |
| 2000 | 28.1% | 16.6% | 10.6% | 25.6% |
| 2009 | 32.9% | 19.4% | 13.2% | 29.5% |

*Note.* Adapted from *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 2. *Total Fall Enrollment in Degree-Granting Institutions by Race/Ethnicity*

| Group | 1976 | | 2008 | | Change 1976-2008 | |
|---|---|---|---|---|---|---|
| | Total | Percent | Total | Percent | Total | Percent |
| All students, total | 10,985,614 | 100% | 19,102,814 | 100% | 8,117,200 | 74% |
| White | 9,076,131 | 83% | 12,088,781 | 63% | 3,012,650 | 33% |
| Total, Under-Represented Races | 1,690,803 | 15% | 6,353,452 | 33% | 4,662,649 | 276% |
| Black | 1,033,025 | 9% | 2,584,478 | 14% | 1,551,453 | 150% |
| Hispanic | 383,790 | 3% | 2,272,888 | 12% | 1,889,098 | 492% |
| Asian/Pacific Islander | 197,878 | 2% | 1,302,797 | 7% | 1,104,919 | 558% |
| American Indian/Alaska Native | 76,110 | 1% | 193,289 | 1% | 117,179 | 154% |
| Nonresident alien | 218,680 | 2% | 660,581 | 3% | 441,901 | 202% |

*Note.* Adapted from *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 3. *Percentage of adults who have at least a bachelors degree*

| | 1999 | | 2004 | | 2007 | | Percentage Point |
|---|---|---|---|---|---|---|---|
| Country | Percent | Rank | Percent | Rank | Percent | Rank | Change |
| Norway | 25.3% | 2 | 29.4% | 2 | 31.9% | 1 | 6.6 |
| United States | 27.5% | 1 | 29.7% | 1 | 30.9% | 2 | 3.4 |
| Netherlands | 20.1% | 3 | 26.9% | 4 | 29.1% | 3 | 9.0 |
| Israel | dna | -- | 29.0% | 3 | 28.3% | 4 | -- |
| Iceland | 17.8% | 7 | 23.5% | 6 | 26.1% | 5 | 8.3 |
| Denmark | 6.6% | 36 | 25.2% | 5 | 25.5% | 6 | 18.9 |
| New Zealand | 13.1% | 16 | 17.6% | 17 | 25.3% | 7 | 12.2 |
| Canada | 19.1% | 4 | 22.2% | 7 | 24.6% | 8 | 5.5 |
| Korea | 16.9% | 9 | 22.0% | 8 | 24.4% | 9 | 7.5 |
| Australia | 17.7% | 8 | 21.9% | 9 | 24.1% | 10 | 6.4 |

*Note.* Adapted from *Education at a Glance* by the Organisation for Economic Co-operation and Development, 2009, Paris, France.

Table 4. *Undergraduate enrollment in US colleges and universities*

| Year | Full-Time | | Part-Time | | Total | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| 1970 | 5,280,000 | 72% | 2,089,000 | 28% | 7,369,000 | 100% |
| 1980 | 6,362,000 | 61% | 4,113,000 | 39% | 10,475,000 | 100% |
| 1990 | 6,976,000 | 58% | 4,983,000 | 42% | 11,959,000 | 100% |
| 2000 | 7,923,000 | 60% | 5,232,000 | 40% | 13,155,000 | 100% |
| 2008 | 10,255,000 | 63% | 6,111,000 | 37% | 16,366,000 | 100% |
| Percent change from 1970: | | 94% | | 193% | | 122% |

*Note:* Adapted from *The Condition of Education 2010* by Susan Aud et al., 2010, Washington, DC: National Center for Education Statistics.

Table 5. *First-Time Full-Time Enrollment Compared to Total Undergraduate Enrollment*

| Year | First-Time Full-Time Enrollment | | | Other Undergraduate Enrollment | | | Total Enrollment | | |
|---|---|---|---|---|---|---|---|---|---|
| | Number | % of Total | % Increase from 1970 | Number | % of Total | % Increase from 1970 | Number | % of Total | % Increase from 1970 |
| 1970 | 1,587,072 | 22% | -- | 5,781,928 | 78% | -- | 7,369,000 | 100% | -- |
| 1980 | 1,749,928 | 17% | 10% | 8,725,072 | 83% | 51% | 10,475,000 | 100% | 42% |
| 1990 | 1,617,118 | 14% | 2% | 10,341,882 | 86% | 79% | 11,959,000 | 100% | 62% |
| 2000 | 1,918,093 | 15% | 21% | 11,236,907 | 85% | 94% | 13,155,000 | 100% | 79% |
| 2008 | 2,427,740 | 15% | 53% | 13,938,260 | 85% | 141% | 16,366,000 | 100% | 122% |

*Note:* Adapted from *The Condition of Education 2010* by Susan Aud et al., 2010, Washington, DC: National Center for Education Statistics and *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 6. *First-Time Freshmen Undergraduate Enrollment in United States Colleges and Universities*

| Year | Full-Time | | Part-Time | | Total | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| 1970 | 1,587,072 | 77% | 476,325 | 23% | 2,063,397 | 100% |
| 1980 | 1,749,928 | 68% | 837,716 | 32% | 2,587,644 | 100% |
| 1990 | 1,617,118 | 72% | 639,506 | 28% | 2,256,624 | 100% |
| 2000 | 1,918,093 | 79% | 509,458 | 21% | 2,427,551 | 100% |
| 2008 | 2,427,740 | 80% | 596,983 | 20% | 3,024,723 | 100% |
| Percent change from 1970: | | 53% | | 25% | | 47% |

*Note:* Adapted from *The Condition of Education 2010* by Susan Aud et al., 2010, Washington, DC: National Center for Education Statistics and *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 7. *Total Undergraduate Enrollment by Type of Institution*

| Year | Four Year | | Two Year | | Total | |
|---|---|---|---|---|---|---|
| | Total | Percent | Total | Percent | Total | Percent |
| 1970 | 5,049,000 | 69% | 2,319,000 | 31% | 7,368,000 | 100% |
| 1980 | 5,949,000 | 57% | 4,526,000 | 43% | 10,475,000 | 100% |
| 1990 | 6,719,000 | 56% | 5,240,000 | 44% | 11,959,000 | 100% |
| 2000 | 7,207,000 | 55% | 5,948,000 | 45% | 13,155,000 | 100% |
| 2010 (est) | 9,613,000 | 57% | 7,201,000 | 43% | 16,814,000 | 100% |
| Percent Change from 1970: | | 90% | | 211% | | 128% |

*Note:* Adapted from *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 8. *Total First-Time Freshmen Fall Enrollment by Type of Institution*

| Year | Four Year | | Two Year | | All Institutions | |
|---|---|---|---|---|---|---|
| | Total | Percent | Total | Percent | Total | Percent |
| 1970 | 1,113,335 | 54% | 950,062 | 46% | 2,063,397 | 100% |
| 1980 | 1,183,332 | 46% | 1,404,312 | 54% | 2,587,644 | 100% |
| 1990 | 1,127,384 | 50% | 1,129,240 | 50% | 2,256,624 | 100% |
| 2000 | 1,340,760 | 55% | 1,086,791 | 45% | 2,427,551 | 100% |
| 2008 | 1,727,419 | 57% | 1,297,304 | 43% | 3,024,723 | 100% |
| Percent Change from 1970: | | 55% | | 37% | | 47% |

*Note:* Adapted from *Digest of Educational Statistics* by Thomas Snyder and Sally Dillow, 2009, Washington, DC: National Center for Education Statistics.

Table 9. *Full-Time Undergraduate Enrollment in US Colleges and Universities*

| Year | Four Year | | Two Year | | Total | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| 1970 | 4,051,000 | 77% | 1,229,000 | 23% | 5,280,000 | 100% |
| 1980 | 4,608,000 | 72% | 1,754,000 | 28% | 6,362,000 | 100% |
| 1990 | 5,092,000 | 73% | 1,884,000 | 27% | 6,976,000 | 100% |
| 2000 | 5,706,000 | 72% | 2,217,000 | 28% | 7,923,000 | 100% |
| 2010 (est) | 7,659,000 | 72% | 2,936,000 | 28% | 10,595,000 | 100% |
| Percent Change from 1970: | | 89% | | 139% | | 101% |

*Note:* Adapted from *The Condition of Education 2010* by Susan Aud et al., 2010, Washington, DC: National Center for Education Statistics.

Table 10. *Part-Time Undergraduate Enrollment in US Colleges and Universities*

| Year | Four Year | | Two Year | | Total | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| 1970 | 998,000 | 48% | 1,090,000 | 52% | 2,088,000 | 100% |
| 1980 | 1,341,000 | 33% | 2,772,000 | 67% | 4,113,000 | 100% |
| 1990 | 1,627,000 | 33% | 3,356,000 | 67% | 4,983,000 | 100% |
| 2000 | 1,501,000 | 29% | 3,731,000 | 71% | 5,232,000 | 100% |
| 2010 (est) | 1,955,000 | 31% | 4,265,000 | 69% | 6,220,000 | 100% |
| Percent Change from 1970: | | 96% | | 291% | | 198% |

*Note:* Adapted from *The Condition of Education 2010* by Susan Aud et al., 2010, Washington, DC: National Center for Education Statistics.

## References


Achieving the Dream. (2008, July). *Test drive: Six states pilot better ways to measure and compare community college performance* (Policy brief). Boston, MA: Golderberger, S. & Gerwin, C.

ACT, Inc. (2009, January). *Fewer students returning to same school for second year of college* (News Release). Iowa.

Ashar, H. & Skenes, R. (1993). Can Tinto's student departure model be applied to nontraditional students? *Adult Education Quarterly*, 43, 90-100.

Attinasi, L. (1989). Getting in: Mexican Americans' perceptions of university attendance and the implications for freshman year persistence. *Journal of Higher Education*, 60(3), 247-277.

Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M. et al. (2010). *The Condition of Education 2010*. NCES-2010028. Washington, DC: National Center for Education Statistics.

Bean, J. & Metzner, B. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Education Research*, 55, 485-540.

Campbell, C. M. & Mislevy, J. (2009, November). *Students' perceptions matter: early signs of undergraduate student retention/attrition.* Paper presented at the meeting of North East Association of Institutional Research.

Carnevale, A., Smith, N. & Strohl, J. (2010). *Projections of Jobs and Education Requirements through 2018.* Washington, DC: Center on Education and the Workforce at Georgetown University.

Comfort, H., Baker, P. & Cairns, L. (2002). *Transitions project: a qualitative study investigating factors which help and hinder learner progression: report on Phases 1 and 2* (Coalville, Leicestershire and Leicester City Learning Partnership).

Completer. (n.d.) In *Glossary of Integrated Postsecondary Education Data System.* Retrieved November 8, 2010 from http://nces.ed.gov/ipeds/glossary//index.asp?id=129.

The Consortium for Student Retention Data Exchange. (2010). *About CSRDE*. Retrieved November 9, 2010 from http://csrde.ou.edu/web/consortium.html.

Elkins, S., Braxton, J. & James, G. (2000). Tinto's separation stage and its influence on first-semester college student persistence. *Research in Higher Education*, 41(2), 251-268.

The Education Trust. (2010). *About the Education Trust.* Retrieved November 9, 2010 from http://www.edtrust.org/dc/about.

Erikson, E.H. (1970). Reflections on the dissent of contemporary youth. *International Journal of Psychoanalysis, 51*, 11-22.

Hsiao, K. P. (1992). *First-generation college students* (ERIC Digest Report No. EDO-JC-00-04). Los Angeles: ERIC Clearinghouse for Junior Colleges. (ERIC Document Reproduction Service No. ED 351 079.)

Jenkins, D., Ellwein, T., Wachen, J., Kerrigan, M., & Cho, S. (2009). *Achieving the Dream Colleges in Pennsylvania and Washington State: Early Progress toward Building a Culture of Evidence.* Retrieved from Achieving the Dream website: http://www.achievingthedream.org/

Kahn, J. & Nauta, M. (2001). Social-cognitive predictors of first-year college persistence: The importance of proximal assessment. *Research in Higher Education*, 42(6), 633-652.

Karp, D., Holmstrom, L., & Gray, P. (1998). Leaving home for college: Expectations for selective reconstruction of self. *Symbolic Interaction*, 21(3), 253–276.

Kemp, W. (2002). Persistence of adult learners in distance education. *American Journal of Distance Education*, 16(2), 65-81.

Kirst, M. & Venezia, A. (Eds.). (2004). *From high school to college: Improving opportunities for success in postsecondary education.* San Francisco, CA: Jossey-Bass.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior* 45, 79–122.

Lumina Foundation for Education. (2009). *Lumina Foundation's Strategic Plan: Goal 2025.* Indianapolis, IN.

MacKinnon-Slaney, F. (1994). The adult persistence in learning model: A road map to counseling services for adult learners. *Journal of Counseling and Development*, 72, 268-275.

Metcalf, H. (1993). *Non-traditional students' experience of higher education: a review of the literature.* (London, Committee of Vice Chancellors and Principals).

National Association of System Heads. (2009). *Access to Success Baseline Metrics: Connecticut State University System*. Retrieved November 9, 2010 from

http://www.NASHONLINE.ORG/System%20Profiles/Connecticut_Bachelors.pdf

.

Nora, A. & Rendon, L. (1990). Determinants of predisposition to transfer among community college students. *Research in Higher Education* 31(3), 235-255.

Organisation for Economic Co-operation and Development. (2009). *Education at a Glance*. Paris, France.

Reyna, R. (2010). *Common College Completion Metrics*. Washington, DC: National Governors Association Chair's Initiative.

Shapiro, A. (2010, September 29). Lumina Foundation's adult degree completion commitment gives millions of recession-battered Americans a "second chance" at earning a degree [Press release]. Retrieved from: http://www.luminafoundation.org/newsroom/news_releases/2010-09-29.html.

Smith, J. & Wertlieb, E. (2005). Do first-year college students' expectations align with their first-year experiences? *NASPA Journal*, 42(2), 153-174.

Snyder, T. & Dillow, S. (2009). *Digest of Education Statistics 2009*. NCES-2010013. Washington, DC: National Center for Education Statistics.

Student Right-to-Know Act. (n.d.). In *Glossary of Integrated Postsecondary Education Data System*. Retrieved November 8, 2010 from http://nces.ed.gov/ipeds/glossary//index.asp?id=625

Transparency by Design. (2010). *About Us*. Retrieved November 9, 2010 from http://www.collegechoicesforadults.org/about.

Tinto, V. (1987). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *Journal of Higher Education,* 59(4), 438-455.

Venezia, A., Kirst, M., & Antonio, A. (2003). *Betraying the college dream: How disconnected K-12 and postsecondary education system undermine student aspirations.* Final Policy Report from Stanford University's Bridge Project.

Voluntary System of Accountability. (2010). *About VSA.* Retrieved November 9, 2010 from http://www.voluntarysystem.org/index.cfm?page=about_cp.

Voluntary System of Accountability. (2010a). *About the College Portrait.* Retrieved November 9, 2010 from http://www.voluntarysystem.org/index.cfm?page=about_cp.

York-Anderson, D. C., & Bowman, S. (1991). Assessing the college knowledge of first-generation and second-generation college students. *Journal of College Student Development* 32(2), 116-122.

# THE DEVELOPMENT OF SOCIAL ACTIVISM DURING COLLEGE

Meredith S. Billings
Senior Research Analyst
Office of Institutional Research & Evaluation
Tufts University

**Abstract**

Researchers at Tufts University are currently conducting a multiyear times-series study on how successful the institution is in instilling the principles of active citizenship in the university community. The longitudinal study focuses on the Tisch Scholars program and its impact on cultivating civic competencies, developing leadership skills, and measuring civic and political engagement in individuals graduating from 2007 to 2009. Each spring, research participants respond to a survey instrument that measures their civic engagement activity levels, attitudes, and beliefs. This paper focuses on the analysis of one civic engagement outcome, the development of social activism during college. Using growth models, the findings indicate that there is a declining linear trend for social activism during college. In addition, demographic characteristics, academic information, and financial aid status are significant predictors for the initial status and growth rate for social activism.

**Key words:** social activism; civic engagement; civic values and beliefs; longitudinal study; growth modeling; campus programs

## Introduction

Currently, there is an overall decreasing trend in civic and political engagement within American society. The youth of today vote less than previous generations and are less knowledgeable about political candidates and causes (Putnam, 1995; Bennett & Rademacher, 1997). In addition, several reports have shown that volunteering among youth with college experience has declined (Lopez et al., 2006; Marcelo, 2007). According to the Civic and Political Health of the Nation surveys, the volunteer rate for youth with college experience declined from 40.9% in 2002 to 36.9% in 2006 (Lopez et al., 2006). This is problematic because a decreasing emphasis on civic values and activities can lead to a disengaged society which would threaten the health and the

strength of the nation. However, higher education can reverse these trends and serve as a vehicle for civic renewal as nearly all professionals and leaders are educated by colleges and universities. Furthermore since there is increasing attendance of all types of citizens at post-secondary institutions, it makes it possible for college and universities to shape the culture of society directly (Colby et al., 2000). Therefore, it is important for higher education to become active participants in citizenship development and emphasize the importance of being engaged in the larger community (Sax, 2000).

Under the leadership of President Lawrence S. Bacow, Tufts University articulated an institutional mission for all students to graduate as committed public citizens and leaders. In 2000, the Jonathan M. Tisch College of Citizenship and Public Service (Tisch College) was established to facilitate and support a wide range of programs that built faculty and student knowledge, skills, and values around civic and political engagement. In the beginning, Tisch College focused on embedding the principles of active citizenship within all fields of study, supporting faculty's civic engagement research, and establishing a set of dynamic community partners. In 2003, university leaders were interested in evaluating Tisch College's progress on its civic engagement initiatives. Therefore, administrators from Tisch College and the Office of Institutional Research & Evaluation (OIR&E) began the Tisch College Evaluation Outcomes Study. The purpose of this study is to examine the links between students' experiences at Tufts University and the development of their civic and political attitudes and activities over time.

This paper focuses on social activism (one civic engagement outcome of the Tisch College Evaluation Outcomes Study) because it is often overlooked in the civic

engagement literature which tends to emphasize community service, service-learning, and political engagement (Lawry, Laurison, & VanAntwerpen, 2006). In addition, there is a need to empower individuals to initiate and sustain positive change for their communities. Therefore, the purpose of this analysis is to help university administrators understand what type of behaviors and demographic characteristics influence social activism in order to create more effective programming and targeted outreach. Therefore, the author addresses two main research questions:

1. How do students vary in their initial levels of social activism?
2. To what extent does students' rate of change in social activism relate to civic values and beliefs, pre-college activity levels, financial aid status, academic information, and demographic characteristics?

## Literature Review

In a large national research study analyzing data from approximately 25,000 college students, Astin (1993) found that the percentage of students who were classified as "social activist" increased from 14% to 25% between freshmen and senior years. This increase of 11% is a net effect and it does not illustrate the movement in both directions on the social activism scale. During this time period, 13% of students studied moved from low to high scores and 5% of students studied moved from high to low scores on the social activism scale. In addition, Astin reported that there was a sizable difference between the percent of freshmen who responded that there were likely to participate in a campus protest and the percent of seniors who had actually participated (7% compared to 25%).

Sax (2000) reported a similar positive growth in students' commitment to social activism during college. She defined commitment to social activism as four life goals: helping others who are in difficulty, influencing social values, participating in community action programs, and influencing the political structure. In all four areas, the percent of students who considered these life goals as "very important" or "essential" increased from a minimum of 5% to a maximum of 18.3% (depending on the particular goal) from freshmen to senior year. Sax also examined students' sense of empowerment, but found there were only slight increases in this civic attitude during college. Empowerment or self-efficacy is important as it has been shown to influence other civic engagement outcomes (Astin, 1993; Sax, 2000; Sax & Astin, 1998). Lastly, Astin studied students' commitment to their communities in high school and during college and found that students who frequently volunteer in high school were twice as likely to be frequent volunteers in college.[1] While volunteering and social activism are different constructs, this is still an important finding because Sax's research provides evidence that pre-college civic engagement levels may affect the development of social activism during college.

While prior research has confirmed that there is a net effect for college on the development of social activism, the more interesting question is what influences these changes. Sax (2000) and Astin (1993) reported that majoring in engineering has a declining effect on students' commitment to social activism. Conversely, enrolling in women's studies and ethnic studies courses has a positive net effect on students' views about social activism and their intentions to engage in social activism in the future (Astin,

[1] Frequent volunteering is defined as volunteering more than three hours per week.

1993; Stake & Hoffmann, 2001). Besides major and course selection, students' race/ethnicity and gender influence the development of civic values (Vogelgesang, 2001). Vogelgesang found in her national study of approximately 20,000 students that freshmen significantly differ on their initial commitment to social activism by race/ethnicity. Black students and Latino students had significantly higher initial means on their commitment to social activism compared to their White counterparts. In addition, White women entered college with a significant higher commitment to social activism than White men. There were also significant interaction effects between race/ethnicity and gender for the development of social activism during college. Asian women and White men were less likely to decrease their commitment to activism during college (compared to their counterparts, Asian men and White women, respectively). In addition, Black men had an overall net increase in their commitment to social activism during college compared to Black women. Lastly, Vogelgesang (2001) and Sax (2000) found independently that participating in community service during college was a strong predictor of increased commitment to social activism.

After examining the relevant literature, the author found that there are several collegiate experiences that have a positive influence on civic values and beliefs. These experiences are majoring in the social sciences (Berger & Milem, 2000); not majoring in the sciences (Rhee & Dey, 1996); discussing political and social issues (Sax, 2000); and receiving financial aid in the form of work-study (U.S. Department of Education, 2000). The author will explore these experiences as potential influences on the development of social activism.

## Methodology

### *Participants*

Each fall, all Tufts University freshmen from the Classes of 2007 to 2009 are invited to take the Tisch College Participant Survey. The survey collects demographic data along with how frequently students participated in civic engagement activities while in high school. Approximately fifteen to twenty percent of the freshmen class elected to participate in the survey each year and respondents to this survey became the population from which the control group samples were drawn. The two control groups are divided by civic engagement activity levels in high school. Those individuals with four hours or more a month in high school are called High School – High Participators (HS High) and those individuals with fewer than four hours a month in high school are called High School – Low Participators (HS Lows). The two control groups (HS High and HS Low) are representative samples of the racial/ethnic, sex, and school affiliation composition of each cohort's first year classmates.

The two control groups are compared to the Citizenship and Public Service Scholars Program (Tisch Scholars). The Tisch Scholars are undergraduate students who are participating in a multi-year civic engagement leadership program. The program consists of an initial civic engagement course, internships with community partners, and independent or group projects to address social issues and needs within the community. Through participating in the Tisch Scholars program, the students build leadership capacity, engage peers and faculty with the values and activities of active citizenship, and create positive change in their communities. All Tisch Scholars from the Classes of 2007 to 2009 are required to participate in the Tisch College Evaluation Outcomes Study and

are not selected to be representative of the student body. Overall, there are 195 participants in the study with 58 Tisch Scholars, 70 HS Highs, and 67 HS Lows for a total of 746 measures. Approximately seven out of the eight participants (87.7%) completed all four surveys. One participant is dropped from the study because the individual only answered the initial survey. Table 1 displays the demographic characteristics of the research participants compared to each entering class.

Table 1. Comparing the demographics and academic information for Tufts University's Classes of 2007 - 2009 to the Tisch Scholars, HS Highs, and HS Lows

| | Tisch Scholars | HS Highs | HS Lows | Class of 2007 | Class of 2008 | Class of 2009 |
|---|---|---|---|---|---|---|
| **N** | 58 | 70 | 67 | 1621 | 1527 | 1536 |
| **Sex** | | | | | | |
| Male | 32.8% | 45.7% | 47.8% | 47.0% | 49.1% | 47.7% |
| **Race/Ethnicity** | | | | | | |
| White | 67.2% | 52.9% | 61.2% | 47.3% | 53.4% | 51.3% |
| Asian | 6.9% | 15.7% | 9.0% | 6.8% | 10.9% | 9.5% |
| Black | 8.6% | 10.0% | 3.0% | 5.5% | 5.3% | 2.9% |
| Latino | 6.9% | 8.6% | 9.0% | 5.7% | 6.1% | 5.3% |
| Multiracial | 1.7% | 1.4% | 0% | 0.2% | 1.8% | 1.4% |
| International | 0% | 1.4% | 1.5% | 1.0% | 2.4% | 2.0% |
| Other | 0% | 0% | 0% | 0.5% | 0.2% | 0.3% |
| Missing/Unknown | 8.6% | 10.0% | 16.4% | 33.1% | 19.9% | 27.3% |
| **Academics** | | | | | | |
| Engineering | 3.4% | 21.4% | 16.4% | 11.0% | 12.0% | 11.4% |
| GPA | 3.51 | 3.35 | 3.46 | 3.30 | 3.34 | 3.34 |

***Data***

Each spring, all research participants complete the annual Civic and Political Activities and Attitudes Survey (CPAAS) for a total of four undergraduate time points. The instrument was developed through compiling questions from eight existing validated civic engagement instruments and soliciting input from national civic engagement

experts.[2] The CPAAS survey questions assess civic and political engagement on campus and within the community as well as the importance and belief in civic values and attitudes. These activity and attitudinal questions also are designed to examine the role that the institution had in developing and influencing active citizenship (Terkla, O'Leary, Wilson, & Diaz, 2007). In addition to the survey data, academic information, financial aid, and demographic characteristics are collected from the Tufts University's Data Warehouse.

***Variables- Level One***

In order to reduce the CPAAS survey items into a smaller number of variables, exploratory factor analysis (EFA) is conducted. The general purpose of EFA is to reduce a large quantity of data into a more manageable set of factors (Meyers, Gamst, & Guarino, 2006). The EFA reveals six factors for students' attitudes and beliefs towards civic engagement. They are labeled as personal efficacy through political action, personal efficacy through community service, social responsibility, cognizance of societal realities, change agency, and acknowledgement of differences. The civic attitudes are entered into the analysis at level one as time-varying covariates, but they are not found to be significant predictors of social activism.

There were two time-varying covariates that are significant predictors of social activism: discussion of social and political issues and the level of participation in community activities. Discussion of issues is a single standardized survey item and the

[2] The survey items were adapted from the following instruments: the Center for Information and Research on Civic Learning and Engagement's (CIRCLE's) Young Citizens'' Survey; Pew's Civic and Political Health of a Nation; the AmeriCorps Baseline Survey; two subscales of the Civic Attitude and Skills Questionnaire (Social Justice attitudes and Diversity attitudes); the Community Service Self-Efficacy Scale; the Public Service Motivation Scale; and the Social Responsibility Inventory.

question asks, "How often do you discuss politics or social issues with your family or friends?" Participants respond on a 4-point Likert scale.[3] Community activities (alpha = 0.748) is created through EFA and is represented by five survey items that are summed and standardized. These five questions measure the number of hours that students participated in community service activities, community-based research, and other community-related events. Appendix 1 displays the survey questions for community activities. Community activities has missing values for 7.1% of its measures and single stochastic regression imputation is employed to resolve these missing values.

The outcome variable, social activism (alpha = 0.692), is constructed by summing and standardizing students' responses to five questions. The five questions are:

*How many hours did you participate in each of the following activities between [specific time period]?*[4]

1. Participate in a protest, march, or demonstration

*How often did you do any of the following between [specific time periods]?*[5]

2. Signed a petition (paper or email) about a political or social issue
3. Wore a button, put a sticker on my car, or placed a sign in front of my house in support of an issue or candidate
4. Not bought something because of the conditions under which the product was made
5. Bought a certain product or service because I like the social or political values of the company that produced it

Social activism has missing values for 5.1% of its measures and single stochastic regression imputation is employed to resolve these missing values.

Two dichotomous time-varying covariates are tested to explore whether they are significant predictors of social activism. The first covariate is whether the participant is currently a registered voter when the individual completed the survey. A value of one

[3] Scale is: 4 = Every day, 3 = Several times a week, 2 = Several times a month, 1 = Never
[4] Survey scale: 6 = More than 120 hours, 5 = 61 – 120 hours, 4 = 26 – 60 hours, 3 = 11 – 25 hours, 2 = 10 hours or less, 1 = None.
[5] Survey scale: 3 = Often, 2 = Seldom, 1 = Never

indicates that the participant is a registered voter while a value of zero indicates that the individual is not a registered voter. The second covariate represents if the participant completed the CPAAS during a year with a national election (Years 2004 and 2008). A value of one indicates that the participant took the survey during an election year and a value of zero indicates that the participant took the survey during a non-election year. The Class of 2007 receives a value of one during their sophomore year, the Class of 2008 receives a value of one during their freshmen year, and the Class of 2009 receives a value of one during their senior year. The two dichotomous time-varying covariates are not significant predictors of social activism and are removed from the model. Table 2 displays the means, standard deviations, and ranges for the significant level one variables for each time point.

Table 2. Means, standard deviations, and ranges for level 1 variables by time

| Variables | Mean | Std. Deviation | Range |
|---|---|---|---|
| **Community Activities** | | | |
| Time 0 | -0.0875 | 0.868 | -1.06 – 2.36 |
| Time 1 | 0.163 | 1.12 | -1.06 – 3.37 |
| Time 2 | -0.112 | 0.920 | -1.06 – 2.56 |
| Time 3 | 0.0350 | 1.06 | -1.06 – 2.97 |
| **Discuss Issues** | | | |
| Time 0 | 0.0659 | 0.989 | -2.34 – 1.30 |
| Time 1 | -0.0679 | 0.984 | -2.34 – 1.30 |
| Time 2 | 0.0376 | 1.01 | -2.34 – 1.30 |
| Time 3 | -0.0374 | 1.02 | -2.34 – 1.30 |
| **Social Activism** | | | |
| Time 0 | 0.204 | 1.078 | -1.64 – 2.77 |
| Time 1 | -0.00896 | 0.983 | -1.64 – 2.37 |
| Time 2 | -0.0921 | 0.995 | -1.64 – 2.73 |
| Time 3 | -0.0774 | 0.963 | -1.64 – 1.97 |

***Variables- Level Two***

At level two, there are several demographic and academic characteristics that the author plans to explore due to significant findings from prior research studies. The variables are sex, race/ethnicity, research group (Scholars, HS High, and HS Low), socioeconomic status (SES), major by discipline (Social Science, Natural Science, Engineering, and Arts & Humanities), and financial aid. Since it is a time-series research design, the author also tests whether there is a cohort effect for social activism. Sex, research group, SES, and cohort status are not significant predictors for social activism and are removed from the model.

Since all of the level two variables are nominal, the author uses dummy codes to distinguish among the groups. For race/ethnicity, participants are divided into five groups (Black, Latino, Asian, White, and Other) and White students are the referent group. The other race/ethnicity category includes individuals who identified as multiracial, who are unknown or missing for their race/ethnicity, or are international students. Given the small sample sizes for multiracial (N = 2) and international students (N = 2), reliable estimates for the regression coefficients cannot be reached and the author decides to collapse these two categories into the other race/ethnicity category. For major, the participants are classified into four broad disciplines (Social Sciences, Natural Sciences, Engineering, and Arts & Humanities) depending on their first major. Arts & Humanities majors are the referent group since prior research has shown that majoring in Social Sciences, Natural Sciences, or Engineering may have an effect on civic engagement outcomes. Lastly for financial aid, the author decides to explore whether receiving financial aid in general is a significant predictor of social activism since there is

only a small number of participants in the study that received work study exclusively (N = 2). The referent group for financial aid is participants who did not receive financial aid. Table 3 displays the mean, standard deviations, and ranges for the significant level two variables.

Table 3. Means, standard deviations, and ranges for level 2 variables

| Variables | Mean | Std. Deviation | Range |
|---|---|---|---|
| **Race/Ethnicity** | | | |
| Black | 0.07 | 0.259 | 0 – 1 |
| Latino | 0.08 | 0.275 | 0 – 1 |
| Asian | 0.11 | 0.311 | 0 – 1 |
| Other | 0.14 | 0.346 | 0 – 1 |
| **Major** | | | |
| Natural Science | 0.14 | 0.352 | 0 – 1 |
| Engineering | 0.13 | 0.335 | 0 – 1 |
| Social Science | 0.45 | 0.498 | 0 – 1 |
| **Financial Aid** | | | |
| Receiving Financial Aid | 0.60 | 0.491 | 0 – 1 |

### *Growth Model*

In order to model change in social activism during college, the author employs a two-level hierarchical model where the first level describes the individual growth trajectory of social activism and the second level uses individual-level variables to predict the initial status and growth (linear change, acceleration, etc.) of social activism. Growth modeling is a form of hierarchical linear models as it treats the repeated measures nested in individuals. It is a more flexible model than repeated measures ANOVA as the number and spacing of measurements can vary across people. In addition, growth modeling allows researchers to model individual growth as a function of person-level and contextual characteristics.

The equations for growth modeling are:

Level 1: $Y_{ij} = \pi_{0i} + \pi_{1i} a_{ti} + \pi_{2i} a_{ti}^2 + ... + \pi_{pi} a_{ti}^p + e_{ti}$

$Y_{ti}$ is the outcome variable for person *i* at time *t*
$\pi_{pi}$ is the growth trajectory parameter *p* for person *i*
$a_{ti}$ is the age at time *t* for person *i*
$e_{ti}$ is the error associated with each person *i*

Level 2: $\pi_{pi} = \beta_{p0} + \sum_{q=1}^{Q_p} \beta_{pq} X_{qi} + r_{pi}$

$\beta_{p0}$ is the effect of $X_{qi}$ on the *p*th growth parameter
$X_{qi}$ is individual-level characteristic of person *i*
$r_{pi}$ is the random error for $\pi_p$

It is essential to test the unconditional model first to correctly specify the individual growth equation and to provide baseline statistics to compare subsequent level two models. In addition to checking the mean growth trajectory and individual variation within the growth trajectory, it is important to explore the reliabilities of initial status and change. The reliabilities of the level 1 coefficients are the ratio between true variance to total observed variance and act as a signal-to-noise indicator. In order to be confident that the variability in the growth parameters are due to true change in the outcome variable and not model error, the reliabilities of the initial status and growth rate should be relatively high. In addition, the reliabilities provide evidence that there are individual differences in initial status and growth rates and that it is appropriate to model each parameter as a function of individual-level variables.

***Level One Models***

The author explores and evaluates three level one models for social activism. Table 4 (on page 16) compares the parameter estimates, reliabilities, and goodness of fit statistics for the three models. The first model is the unconditional linear growth model represented by the following equation: $Y_{ti} = \beta_{00} + \beta_{10} * (time) + r_{0i} + r_{1i} * (time) + e_{ti}$. The fixed effects ($\beta_{00}$, $\beta_{10}$) and the variance components ($r_{0i}$, $r_{1i}$, $e_{ti}$) are significant at $\alpha = 0.05$. It is important to note that the CPAAS requires participants to reflect back on the current academic year in order to answer the survey items. Therefore, initial status (time = 0) is interpreted as the value at the end of the freshmen year and it is not the initial status of the participant as he or she is entering college. The unconditional linear growth model predicts that the average value for social activism at initial status is 0.149 standard deviations and the average growth rate is -0.094 standard deviations. Therefore, the average person has an initial value of 0.149 standard deviations for social activism at the end of freshmen year and the average person's social activism decreases 0.094 standard deviations every subsequent year of college.

To determine whether the rate of change for social activism is different across each year of college, the author examines the unconditional quadratic model (or model 2). Model 2 is represented by the following equation:

$$Y_{ti} = \beta_{00} + \beta_{10} * (time) + \beta_{20} (time^2) + r_{0i} + r_{1i} * (time) + r_{2i} * (time^2) + e_{ti}.$$

While the fixed effects ($\beta_{00}$, $\beta_{10}$, $\beta_{20}$) are significant at $\alpha = 0.05$, there is not significant variation in the growth rate ($p = 0.339$) or in the acceleration ($p > 0.500$) that can be modeled at level two. In addition, there is low reliability for the acceleration coefficient

($\pi_{2i} = 0.041$). Generally, it is recommended to fix the error terms for parameters with reliabilities less than 0.05.

In the third model, the author fixes the error term for acceleration and tests the following hypothesis: $H_0 : \sigma^2 = 0$ and $H_1 : \sigma^2 \neq 0$ to check whether there is significant variation in social activism after controlling for growth rate and acceleration that warrants further modeling at level one. The estimated variance within individuals is $\hat{\sigma}^2 = 0.381$ which is significant at $p < 0.001$. Consequently, the author concludes that time-varying covariates are needed to further explain the variation among time points for social activism. The author tests 10 time-varying covariates and finds two that are significant (involvement in community activities and discussion of political and social issues). The time-varying covariates are entered into the model group centered at level one and the aggregate for each time-varying covariate is added to the model at level two to correctly predict the initial slope. (The time-varying covariates are not significant predictors of growth rate for social activism.) The final model 3 is represented by the following equation:

$$Y_{ti} = \beta_{00} + \beta_{10} * (time) + \beta_{20} (time^2) + \beta_{30} (commact) + \beta_{40} (discuss) + r_{0i} + r_{1i} * (time) + e_{ti}$$

After controlling for time, acceleration, and the two time-varying covariates, there is still significant variation in social activism that warrants further modeling at level one. Unfortunately, there are not any additional time dependent variables available to the author to continue modeling at level one.

Table 4. Parameter estimates, approximate p-values, reliabilities, and goodness of fit statistics for the level one growth models for social activism

| | Model 1 (unconditional linear model) | Model 2 (unconditional quadratic model) | Model 3 (time-varying covariates) |
|---|---|---|---|
| **Fixed effect** | | | |
| Intercept, $\beta_{00}$ | 0.149* | 0.204** | 0.204** |
| Time, $\beta_{10}$ | -0.094*** | -0.265** | -0.268*** |
| Time$^2$, $\beta_{20}$ | | 0.058* | 0.059** |
| Comm Activities, $\beta_{30}$ | | | 0.126* |
| Discuss Issues, $\beta_{40}$ | | | 0.151** |
| **Random effect** | | | |
| Initial status, $r_{0i}$ | 0.742*** | 0.805*** | 0.426*** |
| Growth rate, $r_{1i}$ | 0.037*** | 0.144 | 0.039*** |
| Acceleration, $r_{2i}$ | | 0.004 | |
| Level-1 error, $e_{ti}$ | 0.387*** | 0.371*** | 0.368*** |
| **Reliability** | | | |
| Initial status, $\pi_{0i}$ | 0.727 | 0.695 | 0.617 |
| Growth rate, $\pi_{1i}$ | 0.307 | 0.132 | 0.326 |
| Acceleration, $\pi_{2i}$ | | 0.041 | |
| **Fit statistics** | | | |
| Deviance | 1855.45 | 1850.74 | 1760.04 |
| Parameters | 4 | 7 | 4 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

After examining the reliabilities, fixed effects, random effects, and deviance statistics, model 3 is selected for the final level one model. This conditional level one model predicts the initial status for the average individual's social activism is 0.204 standard deviations $(\hat{\beta}_{00} = 0.204)$, the average growth rate is -0.268 standard deviations $(\hat{\beta}_{10} = -0.268)$, and the average acceleration is 0.059 standard deviations $(\hat{\beta}_{20} = 0.059)$. The fixed effects for model 3 are all significantly different from zero and there are two significant time-varying covariates. The time varying covariates can be interpreted as for every one standard deviation increase in involvement in community activities over and above the group mean (the mean for the individual), social activism is predicted to

increase 0.126 standard deviations ($\hat{\beta}_{30} = 0.126$). For every one standard deviation increase in the frequency of discussing political and social issues with friends and family over and above the mean for the individual, social activism is predicted to increase 0.151 standard deviations ($\hat{\beta}_{40} = 0.151$).

In addition to examining the fixed effects for model 3, the author checks the correlation between individual change and initial status and whether the variance components are significant to warrant further modeling at level two. The estimated correlation between true change and true initial status is -0.488. This means that students who have lower levels of social activism at the end of their freshmen year tend to gain at a faster rate than their counterparts. In order to model person-level differences at level two, it is important to test whether there is significant variation in students' initial status and students' growth rate. (Please note that the quadratic term is fixed since the unconditional quadratic model (model 2) indicated there is not significant variation in students' acceleration to warrant further modeling at level two.) The following hypotheses are tested $H_0 : \tau_{00} = 0$ and $H_1 : \tau_{00} \neq 0$ for the intercept and $H_0 : \tau_{11} = 0$ and $H_1 : \tau_{11} \neq 0$ for the average growth rate. The author rejects the null hypothesis for both and concludes that students significantly differ on social activism at the end of their freshmen year and there is significant variation in students' growth rates to warrant further modeling at level two. To see the estimated parameters for $\hat{\tau}_{00}$ and $\hat{\tau}_{11}$, refer to Table 4.

Lastly, the author calculates the intraclass correlation coefficients (ICC) for initial status and growth rate from the unconditional quadratic model with the fixed acceleration

term. The ICCs provide estimates for how much of the variance in social activism is due to variance between individuals.

$$ICC_{initial\_status} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} = \frac{0.746}{0.746 + 0.381} = 0.662$$

$$ICC_{growth} = \frac{\hat{\tau}_{11}}{\hat{\tau}_{11} + \hat{\sigma}^2} = \frac{0.039}{0.039 + 0.381} = 0.092$$

Therefore, 66.2% of the variance in initial status (at the end of the freshmen year) is due to differences among individuals whereas 9.2% of the variance in growth rate is due to differences among individuals. The ICCs and variance explained from this unconditional model is used to compare subsequent models at level two.

***Level Two Models***

At level two, prior research suggests several demographic and academic characteristics that may explain the person-level differences in social activism. The author groups these level two variables into three themes (pre-college effects, demographic effects, and academic effects) and creates three level two models to test against the baseline model. The pre-college effects model tests whether the research group of the participant (Scholars, HS High, HS Low) is a significant predictor of the initial status or growth rate of social activism. The demographic effects model explores whether sex, race/ethnicity, socioeconomic status (SES), and financial aid are significant predictors of social activism. Lastly, the academic effects model tests whether the discipline that the participant is majoring in is a significant predictor of social activism. Since the pre-college effects model did not find a significant relationship between the research group and social activism, it is not explained further in this analysis.

The two models that significantly predict initial status or growth rate for social activism are the demographic effects model and the academic effects model. The demographic effects model initially contains sex, race/ethnicity, SES, and financial aid. However, sex and socioeconomic status are not significant predictors of social activism and are dropped from the model. While financial aid is also a non-significant predictor of social activism in the demographic effects model, it is retained in the model because the p-value is close to alpha ($\alpha = 0.05$) for predicting the growth rate of social activism. The reliability for the initial status and the reliability for the growth rate are acceptable at 0.618 and 0.305, respectively. Table 5 (on the next page) displays the parameter estimates, variance explained, and goodness of fit statistics for the demographic effects model, academic effects model, and the final model compared to the baseline model.

The significant results of the demographic effects model (Model 4) show that Asian students are 0.385 standard deviations below White students on social activism at the end of their freshmen year ($\hat{\beta}_{03} = -0.385$). There are not significant effects for Black, Latino, and Other students compared to White students on initial status of social activism. There are some interesting results for the growth rate of social activism. On average, Black students increase at a rate of 0.184 standard deviations per year faster than White students on social activism ($\hat{\beta}_{11} = 0.184$). Latino students and Asian students, on average, experience similar growth rates for social activism and they increase at a rate of 0.180 or 0.188 standard deviations per year, respectively, compared to White students ($\hat{\beta}_{12} = 0.180, \hat{\beta}_{13} = 0.188$). The demographic effects model actual explains 42.6% of the variance in initial status and explains 8.0% variance in growth rate. To calculate the

variance explained statistics, the author is comparing the demographics effects model to the unconditional quadratic model with the fixed acceleration rate.

Table 5. Parameter estimates, approximate p-values, variance explained, and goodness of fit statistics for the baseline model, demographic effects model, academic effects model, and final model

| | Model 3 (baseline model) | Model 4 (demographic effects) | Model 5 (academic effects) | Model 6 (final model) |
|---|---|---|---|---|
| **Fixed effect** | | | | |
| Intercept, $\beta_{00}$ | 0.204** | 0.308* | 0.431** | 0.517*** |
| Time, $\beta_{10}$ | -0.268*** | -0.273** | -0.300** | -0.302** |
| Time$^2$, $\beta_{20}$ | 0.059** | 0.058** | 0.059** | 0.058** |
| Comm Activities, $\beta_{30}$ | 0.126* | 0.118* | 0.128* | 0.121* |
| Discuss Issues, $\beta_{40}$ | 0.151** | 0.146** | 0.153** | 0.151** |
| Black, $\beta_{01}$ | | -0.112 | | -0.193 |
| Latino, $\beta_{02}$ | | -0.229 | | -0.229 |
| Asian, $\beta_{03}$ | | -0.385* | | -0.377* |
| Other, $\beta_{04}$ | | -0.181 | | -0.141 |
| Time * Black, $\beta_{11}$ | | 0.184* | | 0.214* |
| Time * Latino, $\beta_{12}$ | | 0.180* | | 0.163* |
| Time * Asian, $\beta_{13}$ | | 0.188* | | 0.172 |
| Time * Other, $\beta_{14}$ | | -0.003 | | -0.007 |
| Financial Aid, $\beta_{05}$ | | -0.025 | | |
| Time * Fin Aid, $\beta_{15}$ | | -0.067 | | -0.084* |
| Natural Science, $\beta_{06}$ | | | -0.313 | -0.266 |
| Engineering, $\beta_{07}$ | | | -0.411* | -0.433* |
| Social Science, $\beta_{08}$ | | | -0.298* | -0.294 |
| Time * Nat Science, $\beta_{16}$ | | | 0.138 | 0.128 |
| Time * Engineers, $\beta_{17}$ | | | 0.085 | 0.130* |
| Time * Soc Science, $\beta_{18}$ | | | 0.001 | 0.011 |
| **Random effect** | | | | |
| Initial status, $r_{0i}$ | 0.426*** | 0.428*** | 0.411*** | 0.409*** |
| Growth rate, $r_{1i}$ | 0.039*** | 0.035*** | 0.037*** | 0.033*** |
| Level-1 error, $e_{ti}$ | 0.368*** | 0.368*** | 0.368*** | 0.368*** |
| **Variance explained** | | | | |
| Intercept | | 42.6% | 44.9% | 45.2% |
| Growth Rate | | 8.0% | 5.1% | 15.4% |
| **Fit statistics** | | | | |
| Deviance | 1760.04 | 1771.56 | 1765.08 | 1776.84 |
| Parameters | 4 | 4 | 4 | 4 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The academic effects model (Model 5) explores whether the discipline of the students' major significantly predicts the initial status or growth rate of social activism. The reliability for initial status and the reliability for growth rate are acceptable at 0.607 and 0.317, respectively. The significant results of the model show that Engineering students are 0.411 standard deviations lower on social activism compared to Arts & Humanities students at the end of the freshmen year ($\hat{\beta}_{07} = -0.411$). Similarly, Social Science majors are 0.298 standard deviations lower on the initial status for social activism compared to Arts & Humanities majors ($\hat{\beta}_{08} = -0.298$). While prior research supports that Engineering students show lower levels of social activism compared to students in other majors, it is a bit surprising that Social Science students are lower on social activism as well. Natural Science majors are not significantly different from Arts & Humanities majors on their initial status for social activism. In addition, there are no significant effects of discipline on the growth rate of social activism. The academic effects model explains 44.9% of the variance in initial status and 5.1% of the variance in growth rate.

The final model (Model 6) combines the demographic effects model and the academic effects model. Since the author wants to avoid misspecification errors due to either the initial status equation ($\pi_0$) or the growth rate equation ($\pi_1$) influencing each other, predictors that are non-significant are removed. Therefore, financial aid is removed from the level 2 equation for initial status, but it is retained in the growth rate equation since it is a significant predictor of growth rate.

The equation for the level one model for the final model is the same as the baseline model:

$$Y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \pi_{2i}(time^2)_{ti} + \pi_{3i}(commact)_{ti} + \pi_{4i}(discuss)_{ti} + e_{ti}$$

The level two model is:

$$\pi_{0i} = \beta_{00} + \beta_{01}(Black) + \beta_{02}(Latino) + \beta_{03}(Asian) + \beta_{04}(Other) + \beta_{06}(NatSci) + \beta_{07}(Engr) + \beta_{08}(SocSci) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(Black) + \beta_{12}(Latino) + \beta_{13}(Asian) + \beta_{14}(Other) + \beta_{15}(FinAid) + \beta_{16}(NatSci) + \beta_{17}(Engr) + \beta_{18}(SocSci) + r_{1i}$$

$$\pi_{2i} = \beta_{20}$$
$$\pi_{3i} = \beta_{30}$$
$$\pi_{4i} = \beta_{40}$$

The combined mixed model is:

$$Y_{ti} = \beta_{00} + \beta_{01}(Black) + \beta_{02}(Latino) + \beta_{03}(Asian) + \beta_{04}(Other) + \beta_{06}(NatSci) + \beta_{07}(Engr) + \beta_{08}(SocSci) + \beta_{10} * (time) + \beta_{11}(Black) * (time) + \beta_{12}(Latino) * (time) + \beta_{13}(Asian) * (time) + \beta_{14}(Other) * (time) + \beta_{15}(FinAid) * (time) + \beta_{16}(NatSci) * (time) + \beta_{17}(Engr) * (time) + \beta_{18}(SocSci) * (time) + \beta_{20}(time^2) + \beta_{30}(commact) + \beta_{40}(discuss) + r_{0i} + r_{1i} * (time) + e_{ti}$$

The reliability for the initial status and the reliability for the growth rate are acceptable at 0.607 and 0.293, respectively. The final model explains 45.2% of the variance in initial status and 15.4% of the variance in growth rate. While the variance explained is the best out of the three level two models, the variance explained is still low especially for the variance explained for growth rate. Therefore, further exploration is needed to explain the individual-level differences in initial status and growth rate in subsequent analyses.

The final model predicts that the initial average value for social activism is 0.517 standard deviations ($\hat{\beta}_{00} = 0.517$) and social activism decreases at an average rate of 0.302 standard deviations per year during college ($\hat{\beta}_{10} = -0.302$). The rate of change for social activism growth increases the declining slope for social activism by 0.058 standard deviations, on average, per year of college ($\hat{\beta}_{20} = 0.058$). The time-varying covariates are interpreted as for every one standard deviation increase in community activities above the group mean (mean for the individual), social activism increases 0.121 standard deviations ($\hat{\beta}_{30} = 0.121$). For every one standard deviation increase in the frequency of discussing political or social issues with friends and family above the mean for the individual, social activism increases 0.151 standard deviations ($\hat{\beta}_{40} = 0.151$).

There are several significant findings of race/ethnicity on social activism for the final model. The results indicate that Asian students, on average, are 0.377 standard deviations lower on social activism compared to White students at the end of their freshmen year ($\hat{\beta}_{03} = -0.377$). Black, Latino, and Other students are not significantly different in the initial status of social activism compared to their White counterparts. There are significant interaction effects between time and race/ethnicity for social activism. On average, Black students increase at a rate of 0.214 standard deviations per year faster than White students for social activism ($\hat{\beta}_{11} = 0.214$). Similarly, Latino students increase at a rate of 0.163 standard deviations per year faster than White students for social activism ($\hat{\beta}_{12} = 0.163$).

Financial aid and students' majors are also significant predictors of social activism. Students who receive financial aid are predicted to decline on social activism at a rate of 0.084 standard deviation per year compared to non-aided students ($\hat{\beta}_{15} = -0.084$). Engineering students, on average, are 0.433 standard deviations lower on social activism compared to Arts & Humanities students at the end of their freshmen year ($\hat{\beta}_{07} = 0.433$). However during college, Engineering majors increase at a rate of 0.130 standard deviations per year faster than Arts & Humanities majors ($\hat{\beta}_{17} = 0.130$). Natural Science and Social Science majors are not significantly different from Arts & Humanities major on their initial status or growth rate for social activism.

## Limitations

There are several limitations to this analysis. The finalized model only explains 45.2% of the variance between individuals' initial status of social activism and a small portion of the variance (15.4%) between individuals' growth rate of social activism. This means that the intraclass correlation coefficient remains high which increases the design effect for the study. Since the effective sample size is the actual sample size divided by the design effect, a high design effect will reduce the sample size to a half or a third of its original size. If there is a small effective sample size, the study may be underpowered and the researcher may fail to find an effect when there is one. Due to this fact, the interpretation of these results should be undertaken with caution. Another limitation to this analysis is the survey instrument changed three years into data collection. In order to have the largest sample size possible, the author only examined survey questions that were on the original survey instrument. Therefore, this analysis may have failed to

capture an important measure that was added to later surveys. Lastly, the initial status for social activism is taken at the end of freshmen year. This is not a true measure of initial status (social activism upon entering college) since the student has already spent a year in college. The college environment may have increased or decreased the original initial status of the participants. Therefore, the author recommends that the findings about the individual differences on initial status are interpreted with caution.

## Discussion

It is not surprising that there is a declining linear trend for social activism in college. While this can be attributed to a number of local reasons, the larger issue is there is an increasing number of college students who are becoming disengaged with their communities and are finding it difficult to initiate and sustain positive change. While is reasonable to believe that students come to college ready to make a difference and may have a high initial level of social activism, they realize through learning more about the issues in the classroom or through experiencing challenges in the field that change does not come quick or easy.

The study suggests several recommendations for university administrators to influence students' commitment to social activism. Due to a significant positive acceleration for a negative growth slope, it is important to target students in the first or second years of college and try to reverse or flatten the declining trend. In addition, there are several actions that increase the development of social activism across all students. These actions are increasing the students' participation in community activities and more frequently discussing political and social issues with friends and family. Therefore, university administrators can plan programs that either involve students within the local

community or provide them an opportunity to discuss social and political issues with their peers. It is important, however, to maintain these activities throughout college and university administrators need to plan accordingly. The author also suggests for college administrators to examine engineering students to determine how this group is increasing at a faster rate during college especially when they entered behind their peers. If there is a specific programmatic reason for this increased growth rate during college, administrators could use this intervention (for a lack of a better term) to help other populations that have a lower initial status on social activism compared to their peers (i.e. Asian students).

Surprisingly, participating in the Tisch Scholars program is not a significant predictor of either initial status or growth rate for social activism. The author has several theories to explain this finding. During the research study, the programmatic components of the Tisch Scholars program were changed and refined several times. Originally, the Scholars program emphasized community service and there was little focus on the development of social activism. However, the Tisch Scholars program has broadened to include more emphasis on various types of civic engagement activities and the participants in the program are more representative of diverse backgrounds and races/ethnicities. Since the participants in the study are more representative of the earlier years of the program, it is possible that the model fails to capture a more supportive and encouraging environment for the development of social activism. In addition, Tisch College was understaffed and its financial resources were reduced during the research study that may have contributed to a non-significant finding.

## Future Research

To the author, the most interesting results from the model are the influence of race/ethnicity and the influence of majoring in Engineering on social activism. Asian students start at a disadvantage by the end of their freshmen year, but continue to grow at the same rate as White students. On the other hand, Latino and Black students who start at relatively the same initial status as White students grow at a faster rate for social activism during college. Therefore, future research is needed to understand how the college environment influences the development of social activism across Black, Latino, White, and Asian students to understand the variations in their initial statuses and growth rates. In addition, Engineering students have a lower value on social activism at the end of their freshmen year, but they grow at a faster rate in each subsequent year of college. Therefore, future research is needed to explore how majoring in engineering influences the development of social activism. Are engineers simply catching up with their peers? Or does majoring in engineering at Tufts University positively influence the development of social activism? These findings can be very useful to university administrators who are interested in cultivating student activists for the benefit and strength of society.

## Appendix 1. Survey Items for Community Activities

Community Activities[6]

How many hours were you involved with this organization or program during [specific time period]?

1. Community service organization
2. Civic issue related conference or seminar

How many hours did you participate in each of the following activities between [specific time period]?

3. Participated in community service
4. Conducted community-based research
5. Attended a meeting of town or city council, school board, or neighborhood association

[6] Scale is: 6 = More than 120 hours, 5 = 61 – 120 hours, 4 = 26 – 60 hours, 3 = 11 – 25 hours, 2 = 10 hours or less, 1 = None

## References


Astin, A. W. (1993). *What Matters in College? Four Critical Years Revisited.* San Francisco: Jossey-Bass.

Bennett, S. E. and Rademacher, E. W. (1997). The 'age of indifference' revisited: Patterns of political interests, media exposure, and knowledge about generation X. In S. C. Craig and S. E. Bennett (Eds.), *After the Bloom: The Politics of Generation X* (pp. 21 – 42). Lanham, MD: Rowman & Littlefield.

Berger, J., & Milem, J. (2000). Organizational behavior in higher education and student outcomes. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 15, pp. 268 – 338). New York: Agathon.

Colby, A., Ehrlich, T., Beaumont, E., Rosner, J., and Stephens, J. (2000). Higher education and the development of civic responsibility. In T. Ehrlich (Ed.), *Civic Responsibility and Higher Education* (pp. xxi – xliii). Phoenix: The American Council on Education and the Oryx Press.

Lawry, S., Laurison, D. L., and VanAntwerpen, J. (2006). *Liberal Education and Civic Engagement: A Project of the Ford Foundation's Knowledge, Creativity, and Freedom Program.* Retrieved May 3, 2010, from http://www.fordfound.org/pdfs/impact/liberal_education_and_civic_engagement.pdf

Lopez, M. H., & Kiesa, A. (2009). What we know about civic engagement among college students. In B. Jacoby and Associates (Eds.), *Civic Engagement in Higher Education* (pp. 31 – 48). San Francisco: Jossey- Bass.

Lopez, M. H., Levine, P. L., Both, D., Kiesa, A., Kirby, E. H., and Marcelo, K. B. (2006). *The 2006 civic and political health of the nation: A detailed look at how youth participate in politics and communities.* College Park, MD: Center for Information & Research on Civic Learning & Engagement.

Marcelo, K. B. (2007). *College experience and volunteering.* College Park, MD: Center for Information & Research on Civic Learning & Engagement.

Meyers, L. S., Gamst, G., Guarino, A. J. (2006). *Applied Multivariate Research.* Thousand Oaks, CA: Sage Publications.

Putnam, R. P. (1995). Tuning in, tuning out: The strange disappearance of social capital in America. *PS: Political Science and Politics,* 28, 664 – 83.

Rhee, B., & Dey, E. L. (1996, October). *Collegiate influences on the civic values of students.* Paper presented at the meeting of the Association for the Study of Higher Education, Memphis, TN.

Sax, L. J. (2000). Citizenship development and the American college student. In T. Ehrlich (Ed.), *Civic Responsibility and Higher Education* (pp. 3 – 18). Phoenix: The American Council on Education and the Oryx Press.

Stake, J., & Hoffmann, F. (2001). Changes in student social attitudes, activism, and personal confidence in higher education: The role of women's studies. *American Educational Research Journal, 38*, 411 – 436.

Terkla, D. G., O'Leary, L. S., Wilson, N. E., & Diaz, A. (2007). Civic engagement assessment: Linking activities to attitudes. *Assessment Update, 19*(4), 1-2, 14-16.

U. S. Department of Education. (2000). *The National Study of the Operation of the Federal Work-Study Program: Summary findings from student and institutional*

*surveys.* (Doc. No. 00-10). Washington, DC: Office of the Under Secretary, Planning and Evaluation Service.

Vogelgesang, L. (2001, April). *The impact of college on the development of civic values: How do race and gender matter?* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

# USING SEM TO DESCRIBE THE INFUSION OF CIVIC ENGAGEMENT INTO THE CAMPUS CULTURE

Meredith S. Billings
Senior Research Analyst

and

Dawn Geronimo Terkla
Associate Provost, Institutional Research, Assessment, & Evaluation

Office of Institutional Research & Evaluation
Tufts University

## Abstract

This study assesses whether Tufts University's campus culture was successful at infusing civic-mindedness in all undergraduates. Civically-minded undergraduates were defined as students who were involved in civic engagement activities as well as those who held civic attitudes and values. A structural equation model was developed and findings revealed that the campus environment had a significant positive impact on civic values and beliefs and a positive indirect effect on civic engagement activities. The model confirmed that there is a supportive campus culture and provides evidence that the institution's mission is successful and verifiable.

**Keywords:** civic engagement; structural equation model; campus culture; institutional mission; civic activities; civic values and beliefs

## Introduction

The development of student citizenship is an important goal of higher education especially as the nation's graduates are faced with solving complex, social problems. Jacoby and Hollander (2009) argue that educating students to become active citizens is not only an essential value of higher education, but fundamental to the future of American democracy and the health of society. In order for higher education to meet these goals, colleges and universities need to institutionalize civic engagement and discuss the importance of active citizenship with their

faculty, staff, and students. One method to institutionalize civic engagement is to emphasize the education of active citizens within the campus mission statement.

At Tufts University, civic engagement is a central tenant of the institutional mission. In fact, the university strives "… to foster an attitude of 'giving back;' an understanding that active citizen participation is essential to freedom and democracy; and a desire to make the world a better place" (Tufts University's Vision Statement, 1994-95). In addition, Tufts University strengthened its commitment to educating public citizens and leaders by establishing the Jonathan M. Tisch College of Citizenship and Public Service (Tisch College) in 2000. The purpose of Tisch College is to foster a culture of active citizenship throughout the university and to build faculty and student knowledge, skills, and values around civic engagement. Initially, Tisch College focused on integrating civic engagement courses into the curriculum, supporting civic engagement research, and developing a set of strong partnerships with community organizations. Currently, Tisch College has clarified its strategy and works with four key constituencies (students, faculty, community partnerships, and alumni) with varying degrees of intensity (Hollister, Mead, & Wilson, 2006).

In an effort to evaluate the civic engagement initiatives, the Office of Institutional Research & Evaluation (OIR&E) along with administrators from Tisch College launched a series of research studies. This paper focuses on one of those studies. The authors collected data regarding undergraduates' civic engagement activities, attitudes, and values to gain a better understanding of how the Tufts environment influenced and shaped the students' development in these areas. The main objectives of the study were to assess the effectiveness of Tufts University's mission of infusing the principles of active citizenship within its students and to

provide empirical evidence that a supportive campus culture can affect civic engagement outcomes. For the purposes of this paper, three main research questions are addressed:

1. How does the campus environment affect the civic attitudes and beliefs of students?
2. How does the campus environment affect the civic engagement activities of students?
3. Does the campus culture have a different impact on male and female students? or on students of color and white students?

## Literature Review

There has been an increasing trend to educate college students to become informed, active citizens for the well-being of their communities. In order for higher education to successfully address this mission, colleges and universities need to infuse the principles of civic-mindedness into the curricular and co-curricular activities of the campus. Jacoby and Hollander (2009) argue that, "civic engagement must be woven into the fabric of the institution if it is to be successful over time" (p. 228). In addition, they offer three campus-based strategies to cultivate and sustain civic engagement: (1) to develop campus-wide infrastructure for civic engagement, (2) to provide access and opportunity for all students regardless of race, ethnicity, social class, religion, politics, and (3) to demonstrate the long-terms effects of civic engagement to the individual and to society. Since institutionalizing service-learning through the development of a campus-wide infrastructure has been successful (Pigza & Troppe, 2003; Furco, 2001; Hollander & Saltmarsh, 2000; Holland, 1997; Bringle & Hatcher, 1996), Jacoby and Hollander proposed that this model can be easily adapted to institutionalizing civic engagement. They further

recommend that the institutional mission, strategic plan, and presidential speeches contain or emphasis the importance of civic engagement. In addition, supporting democratic classrooms, involving students in campus government, creating campus policies that encourage student involvement, and tailoring the approach of student affairs professionals are also other methods to institutionalize civic engagement (Jacoby & Hollander, 2009; Hoffman, 2006).

Hoffman (2006) emphasizes that the campus culture is essential in educating citizen-scholars and argues, "students' perspectives and attitudes are shaped by their entire environment, not just the courses and programs designed to teach them" (p. 15). In addition to the recommendations above, he advises to align campus practices with civic ideals and offers several suggestions such as fostering respect and civility to all, welcoming dissenting viewpoints, and building relationships with external communities. Hamrick (1998) also explains how students discern the symbols embedded in the campus culture as support for institutional values. Faculty and staff need to be thoughtful in the messages that they are sending to their students and how their action and inaction may be perceived. One method to bring awareness and support for institutional values is to intentionally send empowering verbal and symbolic messages to the campus community through mission statements and mottos. However, empowering messages are not enough and Hoffman recommends that colleges and universities display these messages in prominent spaces and educate the campus community on how to incorporate the spirit of these messages into their daily lives (2006).

Kuh (2000) found that institutions that emphasized character development as a priority were more successful in developing the desired impact in their students compared to colleges and universities where this was not a priority. Character development was defined as values that relate to moral, ethical, spiritual, civic, and humanitarian areas. In fact, Kuh states, "at these

[value-orientated] institutions, the environment seemed to matter to character development as much (or almost as much) as did the nature of students' expediencies" (n.p.). This is an important finding because it conflicts with previous research that found where students go to college makes little difference in their development (Pascarella & Terenzini, 1991; Pace, 1990). Moreover, Pascarella, Terenzini, and Pace have found that student effort was the most important influence in how college affects students. At Kuh's value-orientated institutions, however, environmental factors are equally important as student effort. He attributes this unusual finding to the fact that these value-orientated institutions have salient missions that emphasize character development. Therefore if character development is important to institutions, Kuh suggests socializing new faculty, staff, and students to value character development, to align institutional policies and practices with the institutional mission, and to create a campus environment where students can develop to their full potential. In addition, faculty, staff, and students need to develop a shared vision of the ideal student experience, to agree on the purpose of the institution, and to outline the expectations for each member in the campus community. Since part of Kuh's character development included areas that encompass civic engagement development, institutions could use his recommendations to help institutionalize civic engagement and instill the principles of active citizenship throughout the campus community.

While there are several research studies that recommend specific implementation strategies or detail a set of organizational factors to develop the "engaged university" (Hoffman, 2006; Holland, 1997; Bringle & Hatcher, 1996), there are few studies that quantitatively measure the impact that the campus culture actually has on civic engagement outcomes. This study was undertaken to provide empirical evidence that a supportive campus culture significantly affects the civic engagement activities and values of its students.

## Methodology

### *Participants*

The participants in the study include 4,118 seniors from four graduating classes (2005 to 2008) at Tufts University. Tufts University is a private research institution that has four campuses (three in Massachusetts and one in France) and grants graduate, professional, and bachelor's degrees. The main campus is located in Medford/Somerville and houses the two schools (Arts and Sciences and Engineering) that educate undergraduate students. Tufts University attracts academically talented, first time-full time freshmen. The undergraduate student body is equally divided between men and women and approximately two-thirds are from outside of New England. Each year, over 1,300 students graduate with bachelor's degrees and the institution has a consistent four-year graduation rate of 85% ± 2% (Terkla, Topping, Jenkins, & Storm, 2009).

Over half of the participants in this study are female (55.9%) and approximately two-thirds are Caucasian (66.0%) with 12.4% Asian, 6.9% Latino, 6.4% Black, 5.6% International, and 1.5% Multiracial. For the remaining 1.1% of the sample, their race/ethnicity is either missing or unknown. Approximately 7% of participants were transfer students. The sample is equally divided (23.7% to 25.7%) among those who graduated in each year. The majority of the participants received a degree from Arts & Sciences (85.9%) and earned an average GPA of 3.38 ($SD$ = 0.362). Almost half of the participants (47.7%) studied abroad while at Tufts University and 55.2% of the sample indicated that they had participated in community service or civic engagement activities while in college.

### *Data*

The data source for the study is the annual senior survey that is administered to the senior class during their final spring semester. Typically, the senior survey is completed by over 95% of the graduating class and students are queried on a variety of topics: academic advising, curriculum, faculty, post-baccalaureate plans, campus services, and extra-curricular activities. One section of the survey focuses on community service and civic engagement. Specially, 60 items were developed in order to ascertain how undergraduates learned about civic engagement activities, to assess how their civic values and attitudes were shaped by their college experience, and to evaluate their civic engagement activity levels while at Tufts University.

All items were scored on either 4-point or 5-point Likert scales and higher scores indicated more civically-minded individuals. Appendix 1 displays a sample of the civic engagement questions. The survey items were a subset of the Civic and Political Activities and Attitudes Survey (CPAAS). The CPAAS is the primary data source for the Tisch College Outcomes Evaluation Study which is a nine year longitudinal research study examining the link between students' civic engagement activities and their civic and political actions and attitudes throughout college and beyond. The instrument was developed by compiling questions from eight existing validated civic engagement instruments and soliciting input from national experts (Terkla, O'Leary, Wilson, & Diaz, 2007).

### *Data Cleaning*

Prior to data analysis, the data went through several data cleaning steps. Sixty-four participants (1.4%) were deleted from the initial 4,694 seniors because they failed to complete more than half of the civic engagement and community service items. Second, using

Mahalanobis distances, 512 of the participants were identified as multivariate outliers and were removed from the analysis. The outliers were not significantly different from the initial participants based on race/ethnicity, year of graduation, or discipline, but tended to have lower cumulative GPAs[1] and were more likely to be males.[2] In total, the data cleaning process removed 576 of the initial 4,694 participants (12.3%) for a final sample of 4,118. The remaining participants had missing values on some of the survey items in question. The missing values ranged from 0.6% to 8.2% of the cases ($M$ = 3.4%, $SD$ = 2.4%) and single stochastic regression imputation was employed to resolve missing values.

***Structural Equation Model***

To create a structural equation model (SEM), the authors conducted statistical analyses in two parts. In order to reduce the survey questions into a smaller number of variables, exploratory factor analysis (EFA) was conducted on half of the dataset (N = 2043). The general purpose of EFA is to reduce a large quantity of data into a more manageable set of factors (Meyers, Gamst, & Guarino, 2006). The factor structure from the exploratory analysis was tested by confirmatory factor analysis (CFA) on the remaining half of the dataset (N = 2075). CFA is typically employed to determine how well the theoretical factor structure fits the empirical data (Meyers et al., 2006). In the second part of the analysis, the authors used SEM to examine the effects of the latent variables campus environment and students' values and beliefs on the latent outcome variable, civic engagement. SEM is a flexible model that allows researchers to simultaneously test the causal relationships between the variables of interest and examine how well the observed variables represent the underlying latent factors (Kline, 2005).

---

[1] $F(1, 4564) = 19.629$, $p < 0.001$

[2] $\chi^2(2, 4630) = 10.870$, $p = 0.004$

SEM was selected because it has several advantages over regression modeling such as the ability to test the overall model instead of testing individual coefficients, the capacity to model mediating variables rather than solely additive models, the ability to test coefficients across between-subjects groups, and better model visualization due to the graphical interface (Garson, 2009). The structural equation model was analyzed with AMOS 17.0 by maximum likelihood estimation. Assessment of model fit for the SEM was based on four indexes (1) the model chi-square, (2) the Steiger-Lind root mean square error of approximation (RMSEA) with its 90% confidence interval, (3) the Bentler comparative fit (CFI), and (4) the standardized root mean square residual (SRMR). The authors determined that a model with RMSEA $\leq 0.05$, CFI $> 0.95$, and SRMR $< 0.10$ is an excellent fit of the model to the data (Meyers et al., 2006; Kline, 2005).

While theoretical studies explain the importance of the campus culture in developing citizen scholars and empirical research depicts how attending college affects the development of civic engagement outcomes, there is a lack of organizational-level research that quantifies the relationships among the campus environment, students' values and beliefs, and civic engagement activities. After examining the relevant literature, the authors propose using SEM to test the following conceptual model in Figure 1. In addition, the authors examine whether there are differences in the strength and/or direction of the relationships between male and female students and between students of color and white students.

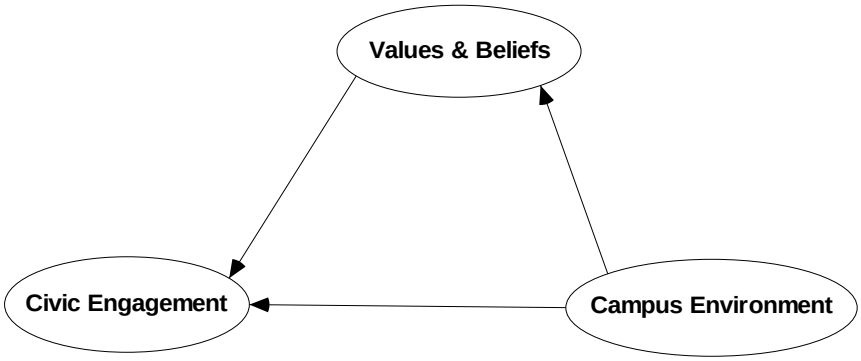



Figure 1. The proposed conceptual model explaining the effects of the campus environment and students' values and beliefs on civic engagement activities

## Findings

### *Factor Analysis*

After reviewing the 60 items on civic engagement activities and attitudes, 15 questions were selected for factor analysis using the principal axis factoring extraction method and a varimax rotation for students' values and beliefs. Two survey items did not load strongly on the factors and were removed. Preliminary EFA revealed three factors for students' values and beliefs that accounted for 63% of the total variance. The three factors were labeled self-efficacy, community connectedness, and leadership ability. Self-efficacy contained five survey items that measured students' perceptions of whether political service and community service are effective ways to create change and whether these activities are an important personal responsibility. Community connectedness also comprised of five items and measured students' increased awareness of issues facing their communities and their interest and responsibilities in serving their communities. Lastly, leadership ability consists of three questions gauging how important it is to the participants to become community leaders or take active roles in specific civic engagement activities or actions.

Similarity, 13 survey items were selected for a factor analysis using the principal axis factoring extraction method and a varimax rotation for campus environment. The campus environment clustered into four factors which were labeled as prevalence of social problems, satisfaction with Tufts, prevalence of unhealthy and risky behaviors, and support for multicultural competency. The four factor solution accounted for 54% of the total variance and three items loaded on factor 1, four items on factors 2 and 3, and two items on factor 4. Prevalence of social problems focused on whether students felt sexual harassment, racism, homophobia, and academic dishonesty were campus problems. Satisfaction with Tufts asked students questions about their overall satisfaction with their undergraduate education, whether their expectations had been met, and how they would rate their academic experience at Tufts. Participants were also asked if given the opportunity to relive their college experience whether they would chose to attend Tufts again. Prevalence of unhealthy and risky behaviors concentrated on whether students felt that alcohol abuse, drug abuse, and eating disorders were campus problems. Lastly, support for multicultural competency evaluated how well the Tufts curriculum or Tufts extracurricular activities prepared students to function in a multicultural society.

The dependent variable, civic engagement, is comprised of two sets of questions. The first set of questions asked students what type of civic engagement activities they participated in at Tufts University. Civic engagement activities were defined as community service, advocacy, political involvement, and community-based research. The second set of questions asked students what type of community and public service activities that they planned to become involved in after graduation. The community and public service activities were defined as volunteering in the community, working for a non-profit organization, participating in service

work through their church, synagogue, or other faith-based organizations, conducting research for social change, making donations to charities or political campaigns, running for elected office, serving on a non-profit board, and attending graduate school in a field related to political or social change. The six items for current civic engagement and the twelve items for future civic engagement were separately summed together to create the two measures for the dependent variable.

CFA suggested several changes to the factor structure. Leadership ability loaded on students' values and beliefs and the outcome variable, civic engagement. In addition, community connectedness loaded on students' values and beliefs and campus environment. Lastly, two factors (prevalence of social problems and prevalence of unhealthy and risky behavior) were dropped from the final structure due to poor loading. The remaining measurement models were confirmed by CFA. Table 1 displays the means, standard deviations, ranges, and Cronbach alphas for the seven observed variables in this study.

Table 1. Mean, Standard Deviations, Ranges, and Cronbach Alphas for the Measurement Model

| Variables | Mean | Std. Deviation | Range | α |
|---|---|---|---|---|
| Self-efficacy | 3.79 | 0.83 | 0 – 5.0 | 0.89 |
| Leadership ability | 2.68 | 0.75 | 1.0 – 4.0 | 0.82 |
| Community connectedness | 3.83 | 0.73 | 0.4 – 5.0 | 0.85 |
| Satisfaction with Tufts | 3.48 | 0.60 | 1.0 – 4.5 | 0.72 |
| Multicultural competency | 3.73 | 0.86 | 0 – 5.0 | 0.57 |
| Current engagement | 0.90 | 1.11 | 0 – 6.0 | 0.56 |
| Future engagement | 4.31 | 2.72 | 0 – 12.0 | 0.78 |

***Revised Structural Equation Model***

The proposed conceptual model (as indicated in Figure 1) was not supported by the data as the path coefficient between campus environment and civic engagement was not statistically

significant ($p = 0.154$). When the relationship between the two variables was dropped, the revised SEM reported the following sufficient goodness-of-fit indices (CFI = 0.989, RMSEA = 0.045, and SRMR = 0.021) and the remaining path coefficients were statistically significant ($p < 0.001$). Although the chi-square test was significant indicating a lack of fit, $\chi^2(10) = 52.496$, $p < 0.001$, Jöreskog and Sörbom (1978) and Bentler (1992) advise against using the chi-square value as the sole predictor of model fit due to chi square's sensitivity to sample size.

Figure 2 represents the revised structural equation model and highlights how the campus environment had a significant positive impact (0.32) on students' civic values and beliefs and a significant positive indirect effect on civic engagement activities of undergraduates (0.24). Students' values and beliefs had a significant direct effect on their level of civic engagement (0.73). In addition, the campus environment was significantly defined and measured by three observed variables: satisfaction with Tufts (0.65), support for multicultural competency (0.63), and community connectedness (0.20). Students' values and beliefs were significantly defined and measured by self-efficacy (0.83), leadership ability (0.37), and community connectedness (0.76). Lastly, the latent dependent variable, civic engagement, was significantly defined and measured by current engagement (0.54), future engagement (0.65), and leadership ability (0.83). Overall, the campus environment explained 10% of the variance in students' values and beliefs and students' values and beliefs explained 54% of the variance in civic engagement as indicated by the $R^2$ statistics. Table 2 displays the unstandardized regression coefficients, standard errors, and p-values for the indicators and latent variables of the revised structural equation model.

Table 2. Unstandardized Regression Coefficients, Standard Errors, and P-Values for the Revised Structural Equation Model

| Parameters | Campus Environment | Values & Beliefs | Civic Engagement |
|---|---|---|---|
| Civic Engagement | | 0.664*** (0.039) | |
| Values & Beliefs | 0.367*** (0.041)[a] | | |
| Self-efficacy | | 1.170*** (0.039) | |
| Leadership ability | | 0.527*** (0.068) | 0.617*** (0.087) |
| Community connectedness | 0.306*** (0.037) | 1.000[b] | |
| Satisfaction with Tufts | 0.928*** (0.077) | | |
| Multicultural competency | 1.000[b] | | |
| Current engagement | | | 1.000[b] |
| Future engagement | | | 1.206*** (0.069) |

*** $p < 0.001$

[a] Standard errors are in parentheses after coefficients

[b] Not tested for statistical significance

In addition to explaining how the campus culture affects the values and beliefs of students and their civic engagement activity levels, the authors tested whether the model is invariant (equivalent) across race/ethnicity and sex. The authors found that there were no significant differences between students of color and white students or between male students and females students with regards to the strength and direction of the relationships among the three latent variables. However, there was a difference in the explanatory power between male and female students. When the path coefficients for males and females were constrained to be equal, civic values and beliefs in male students explained 13% more variance in civic engagement activity compared to female students.

## Discussion

The main research questions in this paper focused on the impact of the campus environment on students' civic attitudes, values, and activities. The results indicate that there is a direct effect of the campus environment on civic attitudes and beliefs and an indirect effect of the campus environment on civic engagement activity levels. The model proposes that there is a

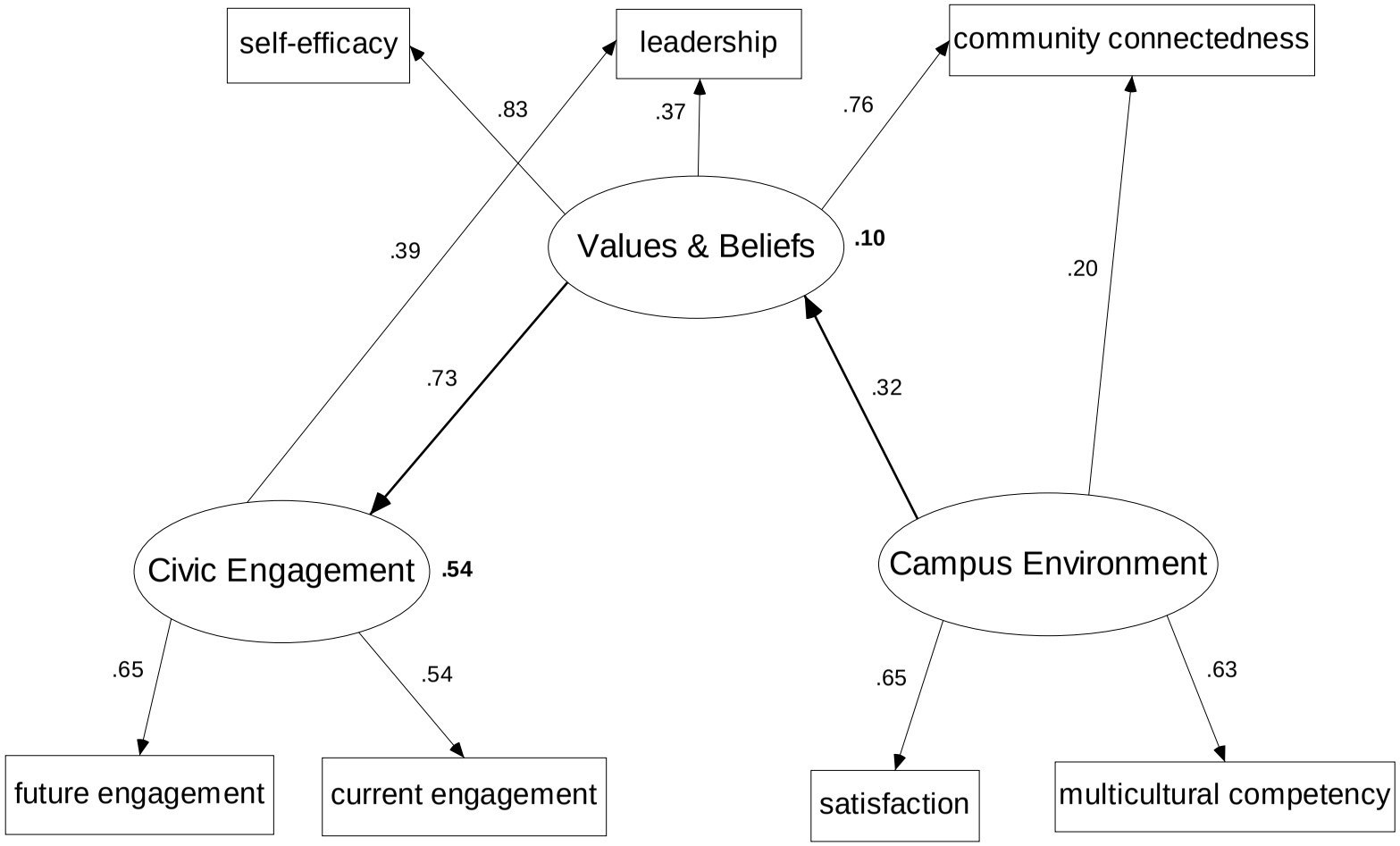



Figure 2. Standardized parameter estimates for the final SEM describing the relationships among campus environment, students' values and beliefs, and civic engagement. *Notes:* Variance explained ($R^2$) is in bold font. Latent variables are in ovals and indicator variables are in rectangles. Estimates shown are significant at $p < 0.001$.

stronger relationship between the campus environment and students' values and beliefs (0.32) compared to the campus environment and civic engagement activities (0.24). However, the strongest relationship in the structural equation model is the correlation between students' civic values and beliefs and civic engagement activities (0.73). Therefore, it is important for higher education institutions whose goal is to develop civically-minded students to focus on fostering supportive campus environments as well as targeting programs and initiatives that will directly affect the civic attitudes and beliefs of its students. Civically-minded individuals are defined as students who are involved in civic engagement activities as well as those who hold civic values and beliefs.[3]

At Tufts University, the institution is committed to developing civically-minded students and actively infuses the principles of active citizenship within the campus community. In the last decade, Tufts founded the Tisch College of Citizenship and Public Service and established the Presidential Award for Citizenship and Public Service for graduating students. Moreover during this period, the President, Provost, Deans, and members of the faculty have emphasized the importance of civic engagement in formal and informal messages to the campus. In addition, administrators, educators, and researchers at Tisch College work with various schools, departments, and student groups to continue to grow the university's capacity for engagement. In recognition of its exemplary commitment to service, Tufts University was selected for the President's Higher Education Community Service Honor Role in 2008 and 2009 (Tufts University named to President's Honor Role for Community Service, 2010). In 2006, the Carnegie Foundation for the Advancement of Teaching chose Tufts University for its new

[3] In this analysis, civic values and beliefs stand for community connectedness, leadership ability, and self-efficacy. However, civic values and beliefs could also represent being informed and responsible citizens, supporting equality and justice for all, understanding complex social problems, appreciating and valuing differences, and encouraging social and political change.

Community Engagement Classification. The award was created to recognize colleges and universities that have institutionalized community engagement in their mission, polices, practices, and culture (Tufts recognized for embracing community engagement, 2007). Lastly, The Princeton Review and Campus Compact selected Tufts University for the book, *Colleges with a Conscious: 81 Greats Schools with Outstanding Community Involvement* (Brand, 2005).

Due to Tufts' awards, recognitions, and institutional action supporting civic engagement, the authors felt confident that the model would confirm a significant relationship between the campus environment and civic attitudes and beliefs and a significant relationship between the campus environment and civic engagement activities. The interesting finding in the study is how the relationship between the campus environment and civic engagement activities is mediated through students' values and beliefs. One possibility is that self-efficacy (belief that one can affect change) influences students' motivation to participate in civic engagement. Without this belief that political and community service makes a difference and can create social change, it is plausible that students will consider their efforts wasted and will be unwilling to devote their limited time to an activity that is unrewarding. Conversely, it is very plausible that a strong self-efficacy may have lasting effects and continue to motivate students to engage in civic engagement activities after graduation. Therefore, university administrators need to design programs that help students increase their self-efficacy and provide them with the necessary tools to initiate positive change. Another possibility is that students need to develop their leadership abilities in order to feel empowered to participate in civic activities. If students do not feel that social issues are important or they do not value being an active participant in social change, they may disengage or avoid civic engagement activities entirely. Thus it is reasonable to posit that

students with strong belief systems that feel they can make a difference will devote their time to civic-minded activities during their undergraduate years and beyond.

## Limitations

A limitation of this research study is that the relationships among the three latent variables may not hold across other colleges and universities since the model used data from a single institution. In fact, the proposed model may only be applicable to institutions that are similar to Tufts University. In addition, colleges and universities that do not foster and provide institutional support for civic engagement may find no significant impact of the campus environment on civic engagement outcomes. This may lead researchers to find different relationships among the three latent constructs for civically engaged institutions and non-civically engaged institutions.

Another limitation of the research study is that the research design did not contain covariates to control for pre-college attitudes and beliefs. Since Tufts University generally attracts civically-minded individuals to their student body, the effect of the campus environment on students' values and beliefs may not be as large as reported since students' initial values may be high when they enter college. Lastly, the civic engagement questions from the senior survey may not fully capture the effect of the campus environment on the development of civic engagement activity, attitudes, and beliefs. It is plausible if the entire CPAAS (and not a subset of the survey instrument) was administered to the same population, the authors would have found a stronger effect.

## Conclusion & Implications for Future Research

The model confirms that there is a supportive campus culture for civic engagement and provides strong empirical evidence that Tufts' institutional mission of service is successful and verifiable. In addition, the model explains how the campus culture can affect students' civic values and beliefs which can in turn affect their level of civic engagement activities. This research is important to institutional researchers, higher education scholars, and university administrators who are interested in the impact of the campus culture on civic engagement outcomes and who intend to use quantitative methods to test whether their institutional missions are reaching all students.

Future research studies should focus on whether this model is generalizable to other institutions. In particular, researchers may discover that the strength of the relationships between the three latent variables vary depending on the type and size of the institution and whether civic engagement has been embedded in the campus culture. Another area of interest is testing whether the institutional mission of civic engagement at Tufts has influenced its staff and faculty. Does working in an institution that is dedicated to active citizenship affect their civic attitudes, beliefs, and activities? How do faculty and staff's actions contribute to the institutional mission of civic engagement?

Since this analysis emphasized the importance of self-efficacy through civic engagement activities, it would be interesting to explore whether there are differences in students' self-efficacy for certain types of current and future civic activities. It is possible that participating in activism and advocacy may require a higher level of self-efficacy than participating in community service or community-based research. In addition, future research should include

evaluating civic engagement programs to document and measure how these programs develop or instill self-efficacy within their students. If evaluators find that some programs are better than others for increasing students' self-efficacy, a further in-depth analysis of these programs may be warranted to understand how they are achieving this goal.

Lastly, graduate and professional students are sometimes overlooked when institutions discuss developing civic engagement outcomes in its students. In an effort to explore civic engagement on the graduate and professional level, the Office of Institutional Research & Evaluation has added several civic engagement questions to its exit and alumni surveys. However, more attention is needed to explore whether graduate and professional students at Tufts University display the same patterns of behavior as the institution's undergraduates. Specifically, do differences exist in the behavior of students who are in disciplines that have embedded civic engagement activities within their graduate programs compared to disciplines where it is not an integral part of the curriculum? How does the development of civic engagement outcomes affect their professional and academic lives? Do graduate students who attend programs that emphasize civic engagement eventually incorporate civic learning into their courses as faculty members? In order to explore these questions, the authors hope to expand the civic engagement sections on the graduate and professional exit and alumni surveys and to conduct future research studies investigating these questions.

## Appendix 1. Sample of Civic Engagement Questions from Senior Survey

Please indicate your level of agreement with each of the statements below:[4]

1. Service to others is valued at Tufts University
2. My Tufts education helped me become more aware of my responsibility to serve my community
3. My Tufts education increased my interest in making change in my community
4. Political service is an effective way to create change
5. Community service is an effective way to create change
6. Being engaged in politics is an important responsibility I have
7. Being involved in making change in my community is an important responsibility I have
8. An undergraduate education should equip students with the skills and knowledge they need to make political and social change

During your time at Tufts, how would you rate your improvement in your understanding of:[5]

9. Problems facing your community?
10. Social problems facing our nation?

How important to you personally is:[6]

11. Helping others who are in difficulty?
12. Participating in a community action program?
13. Becoming a community leader?

[4] Scale is: Strongly agree = 5, Agree = 4, Neutral = 3, Disagree =2, Strongly disagree =1, Not applicable = 0
[5] Scale is: Much stronger = 5, Stronger =4, No change =3, Weaker = 2, Much weaker = 1
[6] Scale is: Essential = 4, Very important = 3, Somewhat important = 2, Not important = 1

## References


Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin, 112*, 400-404.

Brand, H. (2005, June 20). *New Princeton Review book profiles colleges with most outstanding community service programs*, [Press release]. New York: The Princeton Review. Retrieved April 23, 2010 from http://ir.princetonreview.com/releasedetail.cfm?ReleaseID=168268

Bringle, R.G., & Hatcher, J. A. (1996). Implementing service learning in higher education. *Journal of Higher Education, 67*(2), 221-239.

Furco, A. (2001). *Institutionalizing service-learning in higher education: Findings from a three-year study (1997 – 2000).* Paper presented at the Forum on Volunteerism, Service, and Learning in Higher Education, College Park, Maryland.

Garson, G. D. (2009). *Structural equation modeling.* Raleigh, NC: North Carolina State University. Retrieved May 27, 2010 from http://faculty.chass.ncsu.edu/garson/PA765/structur.htm#report

Hamrick, F. A. (1998). Democratic citizenship and student activism. *Journal of College Student Development, 39*(5), 449-460.

Hoffman, D. B. (2006). The campus as civic community: Shaping institutional culture to motivate and empower students as citizens. *Journal of Cognitive Affective Learning, 3*(1), 13-21.

Holland, B. (1997). Analyzing institutional commitment to service: A model of key organizational factors. *Michigan Journal of Community Service Learning, 4*, 30-41.

Hollander, E. & Saltmarsh, J. (2000). The engaged university. *Academe: Bulletin of the American Association of University Professors, 86*(4), 29 – 31.

Hollister, R. M., Mead, M., & Wilson, N. (2006). Infusing active citizenship throughout a research university: The Tisch College of Citizenship and Public Service at Tufts University. *The Metropolitan Universities Journal, 17*(3), 38 – 54.

Jacoby, B., & Hollander, E. (2009). Securing the future of civic engagement in higher education. In B. Jacoby and Associates (Eds.), *Civic Engagement in Higher Education* (pp. 227-248). San Francisco: Jossey-Bass.

Jöreskog, K. G., & Sörbom, D. (1978). *LISREL IV users guide.* Chicago: National Educational Resources.

Kline, R. B. (2005). Principles and practice of structural equation modeling. New York: The Guilford Press.

Kuh, G. (2000). Do environments matter? A comparative analysis of the impression of different types of colleges and universities on character. *Journal of College and Character, 1*(4), Retrieved March 25, 2010, from http://journals.naspa.org/cgi/viewcontent.cgi?article=1277&context=jcc

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research.* Thousand Oaks: Sage Publications.

Pace, C. R. (1990). *The undergraduates: A report of their activities and progress in college in the 1980s.* Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pigza, J.M. & Troppe, M.L. (2003). Developing an infrastructure for service-learning and engagement. In B. Jacoby (Ed.), *Building Partnerships for Service-Learning.* San Francisco: Jossey-Bass.

Terkla, D. G., O'Leary, L. S., Wilson, N. E., & Diaz, A. (2007). Civic engagement assessment: Linking activities to attitudes. *Assessment Update, 19*(4), 1-2, 14-16.

Terkla, D. G., Topping, S. L., Jenkins, M. P., & Storm, S. (2009). *Tufts University Fact Book 2008-2009.* Medford, MA: Tufts University, Office of Institutional Research & Evaluation.

Tufts recognized for embracing community engagement. (2007, February). *Tufts Journal.* Retrieved April 26, 2010, from http://tuftsjournal.tufts.edu/archive/2007/february/briefs/carnegie.shtml

Tufts University named to President's Honor Role for Community Service. (2010, March). *The Active Citizen, 7*(3). Retrieved April 23, 2010, from http://activecitizen.tufts.edu/newsletter/?issue=2010-03

Tufts University's Vision Statement. (1994-1995). Retrieved October 1, 2009, from Tufts University website, Central University Administration: http://www.tufts.edu/home/get_to_know_tufts/vision_statement/