2016 CONFERENCE PROCEEDINGS

TELL-TALE DATA: IR AT THE HEART OF INSTITUTIONAL SUCCESS

NOVEMBER 12-15 2016

HYATT REGENCY ON THE INNER HARBOR

Baltimore, Maryland

CONFERENCE TEAM

President
MARK PALLADINO

Administrative Coordinator
BETH SIMPSON

Program Chair
CHAD MAY

Associate Program Chair
BETSY CARROLL

Exhibitor Coordinator
SALLY FRAZEE

PCW Coordinator
ALLISON WALTERS

Best Paper Coordinator
JENNIFER MAY-TRIFILETTI

Conference Website Coordinator
RICH HETHERINGTON

Evaluation Coordinator
JOEL BLOOM

Member Benefits Coordinators
ELIZABETH CLUNE-KNEUER
MELANIE LARSON

Mobile App Coordinator
KENNETH SMITH

Poster Session Coordinator
KRISZTINA FILEP

Proposal Review Coordinator
J.R. BJERKLIE

Publications Coordinator
MELANIE SULLIVAN

Local Arrangements Chair
SHAMA AKHTAR

AV Coordinator
BETSY CARROLL

Conference Photography
GARY BODEN

Dinner Group Coordinators
JOSHUA BITTINGER
NORA BYE
SHAWN LAROCHE

Interpretive Services
JR BJERKLIE

LAC Coordinators
LYNN CHEN
JAMES DICK
PATRICK KELLEHER

LAC Coordinators (cont.)
DENISE KRALLMAN
RAJIV MALHOTRA
LAURA MASSELL
DANIEL SISCO
LINDA SITES

Registration Coordinator
JACQUELINE ANDREWS

Dear NEAIR Colleagues,

It is with great pleasure that I present to you the *Proceedings* of the 43rd Annual North East Association for Institutional Research Conference "Tell-Tale Data: IR at the Heart of Institutional Success".

The 2016 NEAIR Conference welcomed nearly 400 association members to the Hyatt Regency Baltimore Inner Harbor Hotel from November 12th to the 15th where our friends and colleagues enjoyed the hard work of well over 150 presenters. The Association makes public these *Proceedings* as a means of continuously helping to improve the profession by pushing the field of Institutional Research forward on the shoulders of our dedicated membership and by building a spirit of collegiality and #NEAIRINESS.

First, I wish to thank the entire conference planning team; Chad May - Program Chair, Betsy Carroll - Associate Program Chair, Allison Walters - Pre-Conference Workshop Coordinator, Shama Akhtar - Local Arrangements Chair, Sally Frazee - Exhibitor Coordinator, and Beth Simpson - NEAIR Administrative Coordinator, who together delivered one of the best evaluated NEAIR conferences to date. I'm thrilled to report that the conference evaluation results showed that our members gave this conference record high ratings for both "helping you establish personal and professional contacts" and "addressing the latest developments in institutional research." These two items are central to our mission and are a reflection of the outstanding dedication of this great team to deliver an exceptional conference experience to our membership.

Second, I'd like to express my sincere gratitude to Melanie Sullivan, NEAIR Publications Coordinator, for assembling and disseminating these *Proceedings*, which is no small task. Now that you have returned to your offices, we hope you'll find the *Proceedings* a useful tool as you reflect on the presentations that inspired you at the conference and seek to replicate the good work of your colleagues on your own campuses.

Respectfully,

Mark A Palladino
NEAIR President 2015-16

2015-2016 LEADERSHIP TEAM

Officers:	President	Mark Palladino
	Secretary ('14 -'17)	Jennifer Dunseath
	Treasurer ('13 -'17)	Stephen Sheridan
Steering Committee Members:		
	President-Elect	Ann Marie Senior
	Past President	Bruce Szelest
	Program Chair	Chad May
	Local Arrangements Chair	Shama Akhtar
	Member-At-Large ('13 -'16)	Elizabeth Clune-Kneuer
	Member-At-Large ('13 -'16)	Shannon Lichtinger
	Member-At-Large ('14 -'17)	K. Tracy Barnes
	Member-At-Large ('14 -'17)	Chad Muntz
	Member-At-Large ('15 -'18)	Melanie Larson
	Member-At-Large ('15 -'18)	Tiffany Parker
	Administrative Coordinator (ex-officio)	Beth Simpson

STANDING COMMITTEES

Finance Committee

Chair	Chad Muntz
Chair-elect	TBD 2016
Treasurer	Stephen Sheridan
Member ('14-'17)	Gurvinder Khaneja
Member ('15-'18)	Mary Roetzel
Member ('15-'18)	Laura Qin
Past Chair	Cristi Carson

Grants Committee

Chair	Shannon Lichtinger
Chair-Elect	Tiffany Parker
Immediate Past Chair	Laura Uerling
Grants Reviewer ('14-'17)	Seth Ovadia
Grants Reviewer ('14-'17)	Charis Ng
Grants Reviewer ('14-'17)	June Ragone

Membership

Chair	Elizabeth Clune-Kneuer
Chair-elect	Melanie Larson
Job & News Posting ('14-'17)	Diana Danelian
Mentor ('15-'18)	Daniel Sisco
Newcomer/Mid-Career ('14-'17)	Jennifer May
Newcomer/Mid-Career ('15-'18)	M. Allison Reilly
Social Media ('14-'17)	Claire Goverts
Social Media ('15-'18)	Shawn LaRoche

Nominations (One Year Term)

Chair	Bruce Szelest
Member	Michelle Appel
Member	Michael Duggan
Member	Peter Feigenbaum
Member	Betty Harper
Member - Public	Jason Schweitzer
Member	Jessica Shedd

Professional Development Services (One Year Term)

Chair	**Ann Marie Senior**
Workshop Recruitment	Allison Walters
Local Arrangements	Jane Zeff
Registration	Wendy Weiler

Site Selection (One Year Term)

Chair	**Ann Marie Senior**
Treasurer	Stephen Sheridan
Past LAC or PC	Mark Palladino
Conference Planner	Beth Simpson

Technology

Chair	Tracy Barnes
Chair-elect	TBD 2016
Website ('13-'16)	Joseph Stankovich
Website ('13-'16)	Kenneth Smith
Access/YM ('14-'17)	Melissa Thorpe

NB: Rotation dates refer to Leadership Year start and end date, i.e. '15-'18 means they are listed on the 2015-16 Leadership team and end on the 2017-2018 Leadership team list.

2015-2016 LEADERSHIP TEAM

Program

Chair	**Chad May**
Associate Program Chair	Betsy Carroll
Exhibitor ('13-'16)	Sally Frazee
PCW Coordinator	Allison Walters
Best Paper ('13-'16)	Jennifer May-Trifiletti
Conf Website ('15-'18)	Rich Hetherington
Evaluation ('13-'16)	Joel Bloom
Focus Group & Closing Raffle	Allison Walters '17
	Wendy Weiler '17
	Jane Zeff '17
Member Benefits	Elizabeth Clune-Kneuer
	Melanie Larson
Mobile App ('13-'16)	Kenneth Smith
Poster Session ('15-'18)	Krisztina Filep
Proposal Review ('15-'18)	J.R. Bjerklie
Publications ('15-'18)	Melanie Sullivan

Local Arrangements (One Year Term)

Chair	**Shama Akhtar**
AV Session Checkers	Betsy Carroll
Dinner Group	Joshua Bittinger
	Nora Bye
	Shawn LaRoche
Interpretive Services	J.R. Bjerklie
LAC Coordinators	Lynn Chen
	James Dick
	Patrick Kelleher
	Denise Krallman
	Rajiv Malhotra
	Laura Massell
	Daniel Sisco
	Linda Sites
Photography	Gary Boden
Registration	Jacqueline Andrews

Ad Hoc Committees

Association Software Research

Chair	**Tracy Barnes**
Member	Annemarie Bartlett
Member	Betsy Carroll
Member	Chris Choncek
Member	Beth Simpson

Research

Chair	**Mark Palladino**
Member	Joel Bloom
Member	Alan Sturtz
Member	Beth Simpson

Investment

Chair	**Chad Muntz, Finance Chair**
Member	Stephen Sheridan, Treasurer
Member	Mark Palladino, President
Member	Ann Marie Senior, Incoming President
Member	Bruce Szelest, Past President

NB: Rotation dates refer to Leadership Year start and end date, i.e. '15-'18 means they are listed on the 2015-16 Leadership team and end on the 2017-2018 Leadership team list.

2015-2016 LEADERSHIP TEAM

Conference Proposal Peer Reviewers
J.R. Bjerklie, Proposal Review Coordinator

Shama Akhtar
Lee Allard
Yenny Anderson
Robin Beads
Lisa Belfield
Elina Belyablya
Felice Billups
Joshua Bittinger
Gary Boden
Christina Butler
Marlene Clapp
Elizabeth Clune-Kneuer
Peggye Cohen
Logan Collins
Lauren Conoscenti
Dan Custer
Jerome Dean
Veena Dhankher
Sarah Donnelly
Kate Doria
Erin Dunleavy
Jen Dunseath
Mark Eckstein
Marcia Finch
Gayle Fink
Emily Foley
Sally Frazee
Nora Galambos
Joan Graham
Rachel Groenhout
Karen Heil
Braden Hosch
Christopher Hourigan
May Hser
Kevin Iglesias
Sivakumar Jaganathan
Kathleen Keenan
Heather Kelly
Eric Lanthier
Brian Lashley
Ann Lehman
Christine Lewis-Horton
Shannon Lichtinger
Brad Litchfield
Laura Longo
Pamela MacPherson
Michael Maieli
Laura Massell
Eileen McDonnell
Tom McGuinness
Edmund Melia
Jim Miller
Barbara Moore
Mitchell Nesler
Katherine Ostroot
Allison Ostrowski
Seth Ovadia
Tiffany Parker
George Rezendes
Suzanne Riela
Heather Roscoe
Nathan Rush
Nilay Sapio
Jason Schweitzer
Josephine Seddon
Jessica Sharkness
Jessica Shedd
Ingrid Skadberg
Donna Sohan
Carole Sterritt
Valerie Swain
Carolina Tamara
Melissa Thorpe
Christopher Vinger
Vennessa Walker
Michael Whitcomb
Youyou Zheng

Best Paper Committee
Jennifer May-Trifiletti, Best Paper Coordinator

Best Paper Reviewers TBD

Lee Allard
Marcia Finch
William LeBlanc
Tom McGuinness
Melanie Sullivan
Alexander Yin

NB: Rotation dates refer to Leadership Year start and end date, i.e. '15-'18 means they are listed on the 2015-16 Leadership team and end on the 2017-2018 Leadership team list.

ACKNOWLEDGEMENT

Contained in these pages are the Proceedings of the NEAIR 43rd Annual Conference including contributed papers from NEAIR members.

Conference presentations are accessible within the NEAIR website under the Annual Proceedings section, just a click away. These pages are only accessible to signed in NEAIR members.

Special thanks to Mark Palladino, Chad May, Jennifer May-Trifiletti, Beth Simpson and Tiffany Parker for their contributions and support with all aspects of publications responsibilities during the course of this past year.

Melanie R. Sullivan

2015-2016 NEAIR Publications Coordinator

Providence College

Table of Contents

Observation of a Team-Based Learning Class: Implications for Engagement and Group Dynamics

Joshua D. Bittinger*

For presentation at the North East Association for Institutional Research (NEAIR) Annual Conference, Baltimore, MD, Nov. 2016

*Education Policy, Research & Administration, University of Massachusetts Amherst, Amherst, MA

Observation of a Team-Based Learning Class: Implications for Engagement and Group Dynamics

As college instructors shift away from their roles as sole purveyors of knowledge in favor of more interactive manners to teach material, team-based learning (TBL) has caught the eye of many. TBL was first conceived of by Larry Michaelsen and was used in the context of business education (Wilson-Delfosse, 2012). Since then, other disciplines have bought into the notion, largely in medical- and health-related fields. In theory, adapting a TBL method allows instructors to spend more time assisting students during class. Students spend the majority of their class time actively engaged in working through problems with their peers. This potentially new group work aspect of class adds a layer of responsibility and accountability to coming to class prepared (Michaelsen & Sweet, 2008). Yet, little meaningful research has looked into whether or not TBL classes are working as anticipated.

With the application of a new pedagogy comes the challenge of determining whether or not the newly adopted way of doing things is actually any better than the old way. Studies purporting to evaluate the effectiveness of TBL have largely focused on student engagement as an indicator of pedagogical success (e.g. Chung, Rhee, Baik, & A, 2009; Dunaway, 2005). Students in TBL classes nearly always end up more engaged in their coursework than students in lecture-based courses (Sisk, 2011). However, the same does not hold true for satisfaction rates. Results have indicated a mixed reception for TBL among students (e.g. Abdelkhalek, Hussein, Gibbs, & Hamdy, 2010; Parmele, DeStephen, & Borges, 2009). Grade comparisons have yielded more optimistic results for TBL (e.g. Carmichael, 2009; Chung et al., 2009), yet the question stands for all measures: how should we evaluate the effectiveness of TBL applications?

TBL evaluations focus on both the faculty and student perspectives. From the faculty perspective, studies have shed light on practical applications and operation of TBL (e.g. Searle Haidet, Kelly, Schneider, Seidel, & Richards, 2003; Walters, 2012). Fewer studies explore the student experience. When they do, researchers use short surveys to measure satisfaction and/or levels of engagement (e.g. Chung et al., 2009; Haidet, Morgan, O'Malley, Moran, & Richards, 2004). However, these measures are problematic because they do not capture the entirety of the student experience and measuring engagement levels in TBL classrooms will likely always results in biased results. Missing from these measures is insight into how the groups operate. Additionally, does TBL solve the problem of the bored, distracted student in a lecture hall who has trouble paying attention and staying on task?

In light of this needed information, I directly observed two sections of a single course employing a TBL approach over the course of a semester. My only role in the classroom was to observe, allowing me to focus on what was actually happening in the groups as opposed to also assisting in an instructional manner. In the following sections, I review literature around the development and evaluation of TBL thus far as well as provide detailed context about the course observed. This context is important as it undoubtedly played a role in shaping the student interactions which transpired. I present results focused on several core themes that emerged from my analysis and provide a discussion to put these results into context with the rest of the literature. The following questions guided my inquiry:

How actively engaged are students in a TBL classroom? and

(a) If students are engaged in the classroom, in what ways do they engage in their work?

or

(b) If students are not engaged, how do they disengage from their work?

Literature Review

The literature around TBL is largely based on describing the pedagogical approach or offering tips for instructors wishing to enhance their current application of TBL in their classrooms. I begin this section by reviewing research around the impact of group work in general. Then, I move into a discussion of how TBL is employed in higher education, focusing on recommended practices. This section concludes by presenting previous research on TBL in higher education.

Group Work

Higher education is no stranger to group work, with its use increasing based on pedagogical and employment justifications (Johnson, Johnson, & Smith, 2007; Lejk, Wyvill, & Farrow, 1999; Payne, Monk-Turner, Smith, & Sumter, 2006). Proponents of group work laud its ability to reach students with a variety of learning styles (Hendry et al., 2005). For instance, active learners benefit from experiential activities that can be put into play through group work (Karns, 2006). These learners are disadvantaged in lecture-based courses, which appeal to auditory and visual learners (Davis & Franklin, 2004). In a group, students are able to levy their learning style as their niche and approach work and learning from a stance that best fits their needs (Davis & Franklin, 2004; Hendry et al., 2005).

Group work enhances a variety of skills, including: social development, critical thinking, problem solving, and an appreciation of diversity (MacGregor, Cooper, Smith, & Robinson, 2000). Employers are increasingly looking for students who are able to work effectively in groups, resulting in an increased attention to developing these skills at the postsecondary level (Rundle-Thiele, Bennett, & Dann, 2005). Institutions of higher education meet this demand by including group work in a variety, thanks to its versatility as both a short- and long-term strategy for teaching (Payne et al., 2006).

Prescribed TBL Application

As opposed to traditional lecture courses where students passively absorb information during class, in TBL courses students actively engage with the material during class time. Students prepare for class by completing readings on their own, which are typically concise. By preparing in this manner, students are better able to apply concepts in class and ensure that they know how to use them. Course content is broken down into units which may take multiple class sessions to cover.

At the beginning of each unit, students complete a readiness assessment test (RAT) as individuals (IRAT) and as a group (GRAT). These tests are short and cover the key ideas that students should have grasped from the preparation reading (Michaelsen & Sweet, 2008). The students complete the tests first by themselves and turn in a copy of their responses. They retain a copy so that they can refer to their selected answers when discussing as a group. Groups then complete the same test by discussing answers and coming to a consensus. Based on responses to this group test, instructors may give a concise lecture to review problematic material. The rest of the time is devoted to students working in their teams to solve problems.

Following a TBL approach, instructors are freed up from dispensing information to focus on designing course content and guiding learning (Michaelsen & Sweet, 2008). Compared to other active learning approaches, TBL supporters purport that the technique only requires a single instructor while others require several (Abdelkhalek, Hussein, Gibbs, & Hamdy, 2010; Clark, Nguyen, Bray, & Levine, 2008). This approach helps students move into higher levels of learning, according to Bloom’s taxonomy (Anderson & Krathwohl, 2001). Ideally, students will take care of the first two levels, remembering and understanding facts, before attending class (Walters, 2012). During class, the instructor is then able to focus on the next levels: applying, analyzing, and evaluating. Students also spend time engaging in critical thinking (Hrynchak & Batter, 2012). This requires a shift in expectations on the part of the learner from passive to active, suggested to be a more efficient learning process (Touchet & Coon, 2005).

TBL in Higher Education

Unlike some other pedagogical approaches, TBL relies heavily on small group interaction (Michaelsen & Sweet, 2008). TBL has been seen as attractive to many academic programs due to this mindset shift, but how effective is this approach? Suggested benefits of following TBL include increased student engagement, learning, and satisfaction (Wilson-Delfosse, 2012); however, review of the literature suggests mixed findings.

Studies exploring student engagement found the most consistent results. Searle and colleagues (2003) found that TBL increased the amount students studied outside of class as well as their level of in-class engagement. Several of these authors were also involved in the development of an instrument to measure engagement through observation, which also demonstrated high levels of student engagement in TBL classes (O'Malley, Moran, Haidet, Seidel, Schneider, Morgan, et al., 2003). Similar results have been found in several other studies (Chung et al., 2009; Clark et al., 2008; Dunaway, 2005; Haidet et al., 2004). These results supporting increased levels of engagement by students should not be surprising as students are required to work together in TBL classes (Sisk, 2011).

Studies measuring student satisfaction have found both positive and negative inclinations. Students in a class that utilized a mixture of TBL and problem-based learning generally enjoyed the TBL portion of the class (Abdelkhalek et al., 2010). It is unclear whether the satisfaction with

TBL is influenced by student comparison to problem-based learning. Medical students taking TBL courses over two years were more satisfied with their experiences working in teams but less satisfied with peer evaluation by the end of their second year compared to their initial levels of satisfaction (Parmelee, DeStephen, & Borges, 2009). This finding suggests that as students experience more TBL classes, their satisfaction declines. Compared to students in a lecture-based class, students taking a TBL class reported enjoying class less (Clark et al., 2008). Similarly, students taking a course utilizing TBL did not wish to see the technique used in other classes (Mennenga, 2013). Others found mixed support of student satisfaction (Clark et al., 2008; Searle et al., 2003). Sisk (2011) suggests that student satisfaction should not even be used as an outcome measure because it does not evaluate what a student has learned or whether an approach is effective.

An important link has also been found between student satisfaction and engagement. The more students engage with their peers and the course content, the higher their satisfaction with their experience in a class (Haidet, Schneider, & Onady, 2008). Another study found that as student engagement increased, so did their valuation of team work (Levine, O'Boyle, Haidet, Lynn, Stone, Wolf, et al., 2004). These satisfaction studies have also investigated faculty satisfaction, finding that faculty tend to be quite satisfied with their TBL experiences (Searle et al., 2003; Walters, 2012). No studies were found that set out to explore potential explanations for the mixed satisfaction results.

Survey results from studies exploring engagement and satisfaction shed some light on why students may be reluctant to fully engage in a TBL class. As students take on the first two levels of Bloom's taxonomy on their own before class, some express concerns that they have to teach themselves the material (Mennenga, 2013). This concern makes sense as students have been socialized to be passive learners throughout their schooling (Young, 2009). When students reach TBL classes, often for the first time, they must learn a new way of learning while also grasping course content. Depending on the course, this already challenging process can be exacerbated. Despite their reluctance, faculty members have seen improvements in student attendance, class preparation, and quality of discussion (Thompson et al., 2007).

Outside of engagement and satisfaction, several studies have sought to evaluate the impact of TBL on grades. A study utilizing the individual and group RATs found that students did significantly better on the tests when working as a group as compared to individually (Chung et al., 2009). A similar study found that TBL was particularly effective for more difficult subject matter (Zbheib et al., 2010). Comparing exams scores of students in a TBL class to those in a lecture version of the class, students in the TBL class earned higher grades on the majority of their exams (Carmichael, 2009). Other studies making comparisons between TBL and lecture found no differences between academic performances for the students (Clarke, et al., 2008; Haidet et al., 2004; Mennenga, 2013; Searle et al., 2003). These studies suggest that at minimum, TBL has no effect on grades; however, at best the approach results in an increase in average grades. Despite these findings, the use of grades as a proxy for learning has been challenged (Sisk, 2011).

Another important aspect of TBL that has received little attention is the dynamics of interactions between group members. Much of the work that has investigated these dynamics has explored "social loafing," which occurs when a student fails to contribute to the group in favor of getting a good grade based on the work of that student's group members (Beatty, Haas, & Sciblimpaglia, 1996). Loafers trust that the instructor will only be able to assign grades based on group assignments, unable to distinguish grades for individual students (Su, 2007). Other studies have explored team interactions according to academic performance. One such study found that students worked together best when they had similar academic abilities and if they were high-performers they would work alone if able (Dommeyer, 1986). Low performing students have been found to benefit more from working in groups (Abdelkhalek et al., 2010; Koles, Stolfi, Borges, Nelson, & Parmelee, 2010). This benefit may potentially apply to high-performing students as well (Frame, Cailor, Gryka, Chen, Kiersma, & Sheppard, 2015). A regression analysis did find that low-income and minority students did experience a slightly positive increase in academic outcomes as a result of TBL (Hettler, 2015). A study focusing on seating arrangements states that students intentionally choose their seats, often based on the purpose of the group (Hendrick, Giesen, & Coy, 1974).

The studies reviewed primarily used surveys to collect self-reports of student engagement and satisfaction. When conducting comparisons, authors compared across pedagogical approaches, between individual and group RAT grades, and TBL courses over time. No study was identified in which the researchers directly observed students in a TBL classroom. The closest that any study came was when one of the authors was also the instructor for the course; however, these experiences are potentially biased and incomplete as the instructor was not able to solely observe the students. This purposeful observation is needed, especially in light of reviewed findings above. Directly observing a TBL classroom may shed light on results of student engagement and satisfaction surveys as well as the issue of group dynamics.

Method

The physics course observed took place during the fall semester of 2015 at a large, predominately White, public research university. Six different sections were offered throughout the day with three different instructors. All classes employed a TBL pedagogical approach. The semester of observation was the first semester that each section of the class used TBL. The class was an introductory physics course and was intended to be taken by biology majors and students who were pre-health care professionals. Because of these majors, the course topics were intentionally chosen to be most applicable to this student population and biological examples were incorporated into class exercises. This section will review the context of the course studied and how it diverged from the prescribed TBL model, the sample of students observed, and the methods of data collection and analysis.

Course Context

The course took place in a room specifically designed for TBL classes. The room contained ten tables equipped for team work, with nine chairs located at each table. Each table was allocated three laptops that could be used by students along with access to a dry-erase board and television monitor that laptops could be connected to in order to display a student's screen to the entire table or class. Cameras were also positioned overhead so that a table's dry-erase board could be displayed to the class. The instructor's table was centrally located in the room, and he was able to control what was displayed on the 10 screens located next to each table as well as an additional four monitors located above his station.

Students were randomly assigned into teams of three students, and three teams sat at each table. This assignment broke from the recommendation that groups should contain approximately five students each and intentionally composed to ensure diversity. Before attending class, students completed readings and watched posted lecture videos. Along with the readings and lecture videos, students completed questions based on the material. Students practiced working through the concepts presented during lecture videos during weekly homework exercises; the instructor encouraged collaboration with other students. Students also enrolled in separate lab sections that I did not observe. They completed four exams individually over the semester.

Most classes began with students being given a multiple-choice question to think through on their own as they entered and took their seats. Shortly after beginning class, the instructor asked students to vote on the choice they believed was correct using cards that displayed the letter they chose. Often, there was a moderate to large amount of discrepancy across the room. At this point, the instructor asked students to talk within their teams of three to come to a decision about the answer before calling for a second vote. Typically, after the second vote, the majority of students chose the correct choice. This activity stood apart from typical RAT activities used in TBL classes. Student responses to the questions were not graded either.

Following this activity, students transitioned into working through problems within their teams of three for the remainder of class. During class, the instructor would periodically stop to check in with how students were doing, polling students as to which problems they had completed. When the majority of students had completed a problem, the instructor asked a student to work through the problem on a white board. The cameras were positioned in a way to allow the monitors around the room to display the student's work to the class.

As students worked through problems, the instructor, two graduate teaching assistants, and an undergraduate assistant (hereafter, the graduates and undergraduate are referred to as "assistants") walked around the room answering questions. Very rarely, a small portion of class was dedicated to mini-lectures used to clarify difficult topics. During the latter half of the semester, the instructor asked students to work with their teams at the white boards around the room on all class problems. This allowed instructors and assistants to easily see which problems students were working on as well as where they needed help. This approach also enabled them to guide students down the correct paths earlier instead of following an incorrect solution and losing class time to complete additional problems.

Sample

As mentioned previously, there were six sections of the class that were offered during the observed fall semester. Of these six, two were selected that were offered by the same instructor and were taught consecutively. The first class started at 1:20 in the afternoon and lasted until 2:35. The second class started at 2:40, ending at 3:55. Both classes met twice a week. In order to gain a deeper understanding of how students interacted within their teams over the semester, I observed only two tables during each section. Table 1 contains the number of students at each team as well as the genders of team members. Table letters and team numbers do not reflect actual values assigned to tables or teams. Across both sections, I observed 12 teams and 34 students (split evenly by gender), excluding the student who dropped the class. Five teams were composed of members of a single gender (three man-only and two woman-only), while the remaining seven contained members of each gender.

Table 1. Team compositions by Sections and Tables

	Section 1	**Section 2**
Table A		
Team 1	Woman, Woman, Man	Woman, Woman, Man
Team 2	Woman, Woman, (Man)	Woman, Man, Man
Team 3	Man, Man, Man	Woman, Man, Man
Table B		
Team 1	Man, Man, Man	Woman, Woman, Man
Team 2	Woman, Woman, Man	Man, Man
Team 3	Woman, Woman, Man	Woman, Woman, Woman

Note: A gender placed in parentheses means that the team member did not stay enrolled in the course/section.

Data Collection and Analysis

Classroom observations were conducted over eight weeks during the semester, starting the second week of classes. I negotiated access to the site through conversations with the course instructor, with whom I had previously worked. Prior to my observations, I worked with survey data asking students about their feelings toward group work, approaches to learning, and opinions about TBL. Analysis of these data elicited many questions, whose answers were qualitative in nature. The survey data were only painting a part of the picture, and I sought to better understand the context in which students were responding. To understand this context, I needed to directly observe a TBL class. In this way, I was able to explore the TBL in its natural setting (Nachmias & Nachmias, 1987).

In this setting, I was able to explore a variety of behavior types including: nonverbal, spatial, extralinguistic, and linguistic (Nachmias & Nachmias, 1987). Nonverbal behaviors primarily consisted of body movements and body language exhibited by students. Spatial behaviors focused on where students placed themselves in relation to each other including seating arrangements and the physical proximity while working. Extralinguistic behaviors included how loudly students conversed, whether or not they interrupted each other, and tone of voice. Linguistic behavior accounted for the words students used and structure of conversations.

I observed the two sections one day each of these weeks. During each class session, I sat in the same corner of the room so that I could observe two tables. This allowed me to be close enough to the tables to hear conversations that were above a whisper. At the beginning of each session, I drew a diagram of the seating arrangement of the students at each of the two tables. This was important because students were not assigned seats, so the seating arrangement could vary week-by-week. I recorded extensive observational notes during the class, protecting against recall error or distortions if I would have waited until the class session ended (Nachmias & Nachmias, 1987). Notes were written in a reflexive manner, allowing me to react to observations while in the field. Following class sessions, I journaled to reflect on what I observed during class and begin preliminary analyses.

Accurately observing and recording everything that took place would have been impossible, so I used a combination of induction and deduction to guide my observations (Nachmias & Nachmias, 1987). My first two observations were conducted inductively. During these sessions, I recorded as much as I could, since I did not know what patterns would emerge (Krathwohl, 2009). After these session, I reviewed my notes and focused my observation on interactions within and between teams, as well as how engaged or disengaged students appeared to be with the class activities. When noting engagement, I looked for students actively working through the assigned problems, participating when the instructor polled the class, paying attention to other students when the instructor asked them to work out the problem on a board for the class, and on-topic discussions with other students. Students were noted as being disengaged when they strayed from the above actions.

Observation notes were recorded for tables, teams, and individual students based on the seating diagram that I drew each class session. I did not know the identities of any students, therefore observations had to be made based on phenotypic descriptions so that they could be analyzed over the semester. All observation notes as well as seating diagrams were stored in NVivo 11, which was used to streamline and organize analysis. This aided in analyzing interaction patterns within and between teams. Observations were coded using a priori and inductive, emergent themes (Guest, MacQueen, & Namey, 2012). During analysis, the constant comparative method was employed which required continual reviewing of codes assigned to text to ensure that the observations matched their coded themes (Glaser & Strauss, 1967). My field notes were initially coded based on concepts they represented (Krathwohl, 2009). Comparisons within and between concepts resulted in themes broken into multiple codes to more accurately reflect nuances that were not initially anticipated.

Limitations

Before presenting the findings of my study, a few limitations should be noted. First, the study was conducted by a single researcher. While this allowed students to more easily get used to me presence in the classroom, I am unable to triangulate my notes and interpretations with someone else's. The inclusion of another researcher into the project would help establish credibility and confirmability of the findings I present below (Shenton, 2004). Coming to a consensus with another would demonstrate that the findings are not just based on my own flawed perception and that what I present is actually what occurred.

Second, I rely on phenotypic observations to classify students as either a man or woman. This may be problematic if students do not identify as either. Their non-conforming identities may or may not be the explanation for some group dynamics that I observed. Finally, the course that I observed did not follow all of the prescribed guidelines laid out by TBL experts. This divergence poses a threat to the transferability of the findings if others are dealing with a more traditional TBL approach. However, I have taken care to provide as much detail as possible so that others are able to understand the context in which my findings are grounded (Shenton, 2004).

Findings

The identified main themes are organized under two umbrella categories: group dynamics and student engagement. Each category contains several themes and are discussed below.

Group Dynamics

This section highlights how students worked together within and between teams. Teams are identified by Table letter and Section and Team numbers, as shown in Table 1 above. Gender composition is also noted in parentheses. For example, for Team 1 at Table A in Section 1, the composition is indicated by: (WWM). Student interactions appeared to be influenced by gender, including topics discussed, help-seeking behaviors, and seating locations.

Within-team interactions.

During each class section that I observed, one table of students worked primarily within their teams of three to complete problems while the other table's team boundaries often broke down. Coincidently, students at each Table A worked most cohesively and consistently as teams. In Section 1, Teams 1 (WWM) and 3 (MMM) worked best together, with Team 1 (WWM) working in the fashion expected of students in a TBL classroom. In Section 2, all teams at Table A worked cohesively.

Over the observation period, the all-men team at Table A in Section 1 was the only team whose members attended every class session. When working through problems, this team kept mostly quiet, choosing to work through the problems individually; however, the team members regularly checked in with each other to make sure that each student was able to work through the problem, check answers, and share approaches. When students were polled about their progress, this team was typically one of the teams who was the furthest along. Members always indicated working on the same problem. This team tended to not initiate communication with the other teams at its table, and each student used a separate laptop. No apparent team leader emerged. When the instructor asked students to complete their work on the white boards, this team worked efficiently through problems. Doing so increased their within-team communication as they were all working on a common solution instead of on three separate ones, which contributed to their efficiency.

Also at Table A in Section 1 was the woman, woman, man team which did not lose its man team member early in the semester. This team worked less efficiently than their all-men counterparts. The two women in the team attended every session, while the man missed one. Communication within this team occurred more frequently than the all-men team; students tended to talk through the steps of the problem with each other as opposed to completing the work primarily individually. As opposed to the all-men team, students in this team tended to work off of the same laptop which was typically one located at the table instead of a personal one. Students in this team did not work through the problems as quickly as the all-men team, tending to be one or two problems behind. Discussion within this team tended to be led by the woman student who sat in the middle of the three-person team. Members of this group ended class by exchanging pleasantries such as "See you Monday" or "Have a good weekend."

During Section 2, teams at Table A functioned much more similarly. Teams commonly discussed approaches to each problem with their team members as they were working. Frequently, teams worked off of a single laptop, using the ones provided at each table. When additional laptops were used, students were reading directions to themselves or recording work that they had completed with their teams. When tasked with working at the white boards, all teams worked through the problems as requested, only returning to their table once finished, to check notes, or to record the process for later reference. Communication typically occurred within teams. Teams worked at a pace on par with the class average, with all teams tending to be within a problem of each other.

Between-team interactions.

While the previously mentioned teams at each Table A did interact with other students at their tables, such actions were not the norm. Table B of each section, however, functioned due to interactions across team lines. At each of these tables, some students resisted constraining their interactions to within their teams while others engaged with students in neighboring teams due to absence or disinterest of team members.

In general, there were two primary ways in which students communicated across team lines: as a table and as individuals. At no point did either Table A choose to work through a problem as a table; however, this behavior occurred several times at each Table B. Working as a table occurred early in the semester when students knew little of each other. As the semester progressed, students who were disengaged were systematically excluded from these discussions. While seated, students excluded others by sitting closer together while the excluded student(s) sat on their own. When working on the white boards, the excluded students would either remain seated at their tables or stand near a cluster of students but not contribute toward the effort of the group. Some excluded students self-selected to be excluded by removing themselves from their teams.

When students communicated between teams individually, student gender appeared to influence the conversations. These conversations were aided by proximity – students would choose to engage with other students who sat near themselves. Two purposes predominated: asking for help and talking about shared interests. Early in the semester, students were more apt to talk across teams to get to know one another. Students sought help across teams only once or twice; however, students regularly talked about shared interests and typically with the same peers. More detailed analyses of both of these conversation types will be discussed subsequently in terms of gender.

Same-gender interactions.

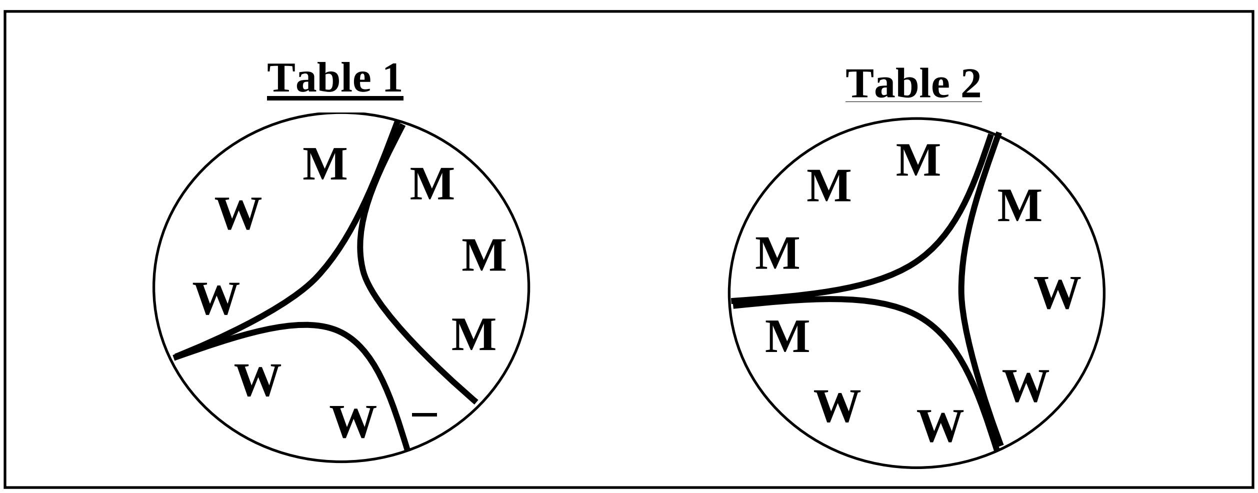
When students communicated across teams, they most frequently engaged with students of the same gender. This pattern held true for help-seeking but particularly shared interest interactions.

Every team contained at least two members of one gender. In mixed-gender teams, women tended to outnumber men. While students readily spoke with and sat next to others of the same gender, teams of only a single gender did not fare well. The one exception was the all-man team at Table A in Section 1. These three men worked together cohesively and efficiently throughout the semester. For the other single-gender teams, students either assimilated with other teams or remained silent. The prime example of the assimilation pattern was the all-men team at Table B in Section 1. These three men worked as a team on occasion, but more often than not would form a larger cluster with the other men at the table. Their attendance was inconsistent, and at least one student very much disliked the TBL approach. During one class session, he approached me and said, “I hate TBL.” At times, this cluster of men worked efficiently through problems, yet they often drifted from discussing class material in favor of other topics such as TV shows or college sports teams. Other times, the cluster seemed disjointed as different men worked on different problems. On multiple occasions, the man who hated TBL would leave the cluster to ask an assistant a question, which the two would subsequently work out on a white board.

Neither team of only women worked well together. In both, there was a disengaged student who did not contribute to discussion and refused to participate in class activities. Likely because of this, the women who wished to work on the problems would try to work with a neighboring team. These collaborations were not long-lasting as they seemed to be nonexistent every other week. Even when attempting to assimilate with a neighboring team, the students did not contribute much. They spoke minimally and appeared to only write down things that the neighboring team said. This was similar to the interactions between the all-man team and the two external men when they all formed a cluster as discussed above.

When students communicated with their peers in other teams, they commonly did so along gender lines. This was facilitated by their chosen seats, as members within a team would arrange themselves so that the outside members (as opposed to the student seated in the middle of the other two students) sat next to students of the same gender (see Figure 1). These outside members of teams would then serve as links to the other teams at the table. When a team had a question, a student sitting on the outside of the team would ask for help from a neighboring student of the same gender. Before dropping the course, the man from the Team 2 of Section 1 (WWM) exhibited this behavior. When working in his team, he only listened and sat at a distance from his teammates. When he had a question, he would turn to a neighboring man (from the all-man team) for help or would request the help of an assistant. He played a passive role on his team, only taking in information.

Figure 1. Seating arrangements for two tables of Section 1



Different-gender interactions.

Interactions between students of different genders happened almost exclusively within the context of their teams. The level of interaction varied by team, with some teams talking through all steps of problems while others only talked when they encountered problems. At Table A in each section, the mixed-gender teams worked as cohesive teams and generally communicated with each other frequently. Their counterparts at Table B were not so consistent. However, these teams interacted more frequently than either of the woman-only teams.

When different-gendered students interacted with students outside of their team, they were often seeking help. This occurred most often when a woman student would ask a man student to explain the process of solving a problem to her. An incident such as this occurred early in the semester at Table B of Section 1 when a woman asked one of the men in the all-man team for assistance. This interaction was not positive and the man sounded demeaning to the woman, having little patience for explaining concepts he understood. The woman did not ask him another question the rest of the semester. Interactions based on shared interests were exceedingly rare, with my only recorded observation occurring when a man asked a woman if she dyed her hair and whether she was in his lab section.

Student Engagement

In general, students from each section were likely to be engaged in their work at any given moment. However, there were certain students whose engagement levels and patterns differed significantly from others. Those differing behaviors are presented here at two levels. The first level, disengagement from class activities, includes different ways that students strayed from the task-at-hand. The second level, removal from one's team, describes extreme measures taken by a few students to physically move themselves away from their teams.

Disengagement from class activities.

Disengagement took a number of forms, but most frequently occurred with the assistance of technology. The two primary categories of disengagement were interpersonal and technology-assisted. Interpersonal disengagement occurred when students' conversations strayed from their class activities. After exams, these conversations revolved around gathering thoughts from their peers about the perceived difficulty of the assessments. During these classes, the instructor had to continuously ask the students to cease conversations while he or other students were talking to the class. Section 1 was particularly persistent in their desire to talk about the exams, requiring multiple requests from the instructor. Section 2 was more responsive to instructor requests. In both sections, the conversations about exams permeated into the time students were supposed to work with their teams to solve problems.

Other interpersonal disengagement occurred mainly between students of the same gender, as mentioned previously. At the beginning of the semester, these conversations were more focused on gathering information about their peers such as majors and what classes they were taking. As the semester progressed, these conversations turned toward shared interests such as popular television shows and football. Similar to conversations pertaining to exams, students frequently continued discussions while the instructor or other students talked to the class. At no point in the semester did these conversations stop happening, so the instructor requesting students to quiet down was a regular occurrence.

Technology-assisted disengagement took place when students used laptops (both provided and personal) for purposes other than working on the assigned problems or when they were observed sending text messages. For the most part, students refrained from using the provided laptops for non-course-related activities. Instead, students used their personal laptops for activities such as working on homework for another class, checking e-mail or Facebook, and managing their fantasy football teams. Some students alternated between working on the assigned problems and the above activities.

As the semester progressed, the frequency of students sending text messages increased. Toward the end of the semester, I observed one student enter class late, sending text messages on the way to his seat, then continue to send messages the rest of the class session. He did not stop when attempting to participate in class. At one point the instructor asked for a student to explain an answer to the class and the texting student raised his hand while continuing to text with his other hand. Another student was called on, so he started to send messages with both hands again. Shortly thereafter, the instructor asked for another volunteer, at which point he put a hand back up. The most flagrant instances of students sending messages were observed when the instructor was talking.

Removal from one's team.

Across the semester, a number of students removed themselves from their teams. Here, removal occurred when a student either physically distanced him/herself from or chose not to work in conjunction with his/her team member(s). Removal sometimes worked in conjunction with disengagement as discussed previously.

Removal most commonly occurred when students physically moved their chairs away from their team members. This behavior was primarily exhibited by two students (one woman and one man). The man was a member of the two-man team at Table 2 in Section 2. He would often push himself away from the table so that his chair was several feet away from his team and table. He did this when his teammate was present or absent. This removal typically occurred after he had finished working on a problem. The woman was part of Team 2 at Table B of Section 1 (WWM). Instead of pushing back from the table, she would sit away from other students. At times, she would leave a chair or two between herself and her teammates. When the instructor would come near the table, she would move closer to her team members. While sitting apart from her team, she spent most classes playing on her tablet or completing homework for other classes.

Discussion

My investigation was framed around better understanding student engagement in ways not possible through survey administrations alone. Unlike some other studies, I do not compare my results to a lecture course. The reason for not doing so is two-fold. First, there was no lecture-based section of this course offered during the observed fall semester. Second, doing so would likely result in similar findings about engagement as previous studies (Sisk, 2011). However, I draw a few comparisons to lecture-based courses in general. I also did not explore grades as an outcome for this course as I was unable to link student grades back to the observed teams. This was not problematic as, overall, students did very well in the course, resulting in little grade variation. Instead of using grades as a proxy for a student's level of understanding of course content, I was able to utilize how quickly students progressed through the problems in comparison to their peers when the instructor polled the class to determine which problems students were working on. When students began working at their white boards as teams, I was able to use their completion of steps and subsequent returns to their seats as an additional indicator.

As stated previously, some proponents of TBL suggest that only a single instructor is required (Abdelkhalek et al., 2010; Clark et al., 2008). They suggest that other forms of active pedagogy require several instructors, in comparison. However, my observation raises questions about this declaration. In the course I observed, there were four people actively assisting students across thirty different teams. Even with this many people, students often found themselves waiting for assistance for several minutes. The instructor and assistants were engaged with students during the entire session. When the teams began working everything out on their white boards, the instructor and assistants became even busier because they were able to see students making mistakes earlier. I find it hard to believe that a single instructor could effectively manage and guide all of the teams alone.

As prescribed, TBL teams ought to be assigned deliberately with attention paid to ensuring a diverse representation of students, despite evidence suggesting that students who self-select group members report higher rates of satisfaction, commitment, and cognitive learning (Myers, 2012). This recommendation holds true outside of TBL assignment (Lighfner, Bober, & Willi, 2007). Randomly assigning students to groups may negatively impact minority students (Hinds, Carley, Krackhardt, & Wholey, 2000), especially considering that the course was offered at a predominately White institution. Initially, I noted whether or not students were White (based on phenotypic observations), but this proved problematic. Additionally, group dynamics did not appear to have been affected by student race or ethnicity. Students in the course tended to be high performing students; many were pursuing a pre-med degree. Given the lack of variation in academic ability and race/ethnicity, gender appeared to be the key form of diversity in the teams. The overall lack of diversity suggests that random assignment may not have been problematic in this case. Initially, groups were supposed to be reassigned halfway through so that students could work with new students. However, this did not come to fruition, which aligns with best practices (Delucchi, 2006).

My findings around group dynamics lead to several questions that should be pursued through further research. Team 3 of Table A in Section 1 (MMM) worked more efficiently than the other three-person team (WWM) at their table. However, is efficiency the desired outcome for TBL groups? The all-man team did not work collaboratively during the first half of the semester. Instead, they worked individually on each problem during class. Only once they had finished a problem individually would they communicate with each other to check to make sure that they all had the correct answer. This behavior likely benefited their performance on exams, because exams were completed individually. However, TBL groups would ideally function more similarly to the other team (WWM). This second team talked through each aspect of the problem. When there was no consensus about how to proceed, each student shared an approach with the team. After this sharing, the team would try out an approach or seeking help from the instructor or assistants. These conversations delayed their ability to proceed through the questions in a timely manner and rarely completed all of the problems assigned. Yet, this back-and-forth process and group approach to problem-solving is one of the skills that is being fostered through a TBL approach. The difference in efficiency could also be due to other factors, such as how prepared students from each team were when they came to class.

The all-man group did begin working more collaboratively during the second part of the semester when the instructor asked students to show all of their work on the white boards around the classroom. At this point, the men broke with their individualistic approach to problem solving. Instead, they took turns working out problems on the board and began to discuss the steps they needed to take to solve each problem. Having students show their work on a white board pushed these men out of their comfort zone and seemingly facilitated the development of group problem-solving skills. Their rate of problem completion decreased slightly, but they were still able to complete all problems during class and rarely required assistance. Such an intervention by the instructor proved particularly useful in this scenario. This movement to the white boards also allowed the instructor to keep students on track in other ways, enhancing his ability to guide students (Payne et al., 2006).

The role that gender plays in TBL group dynamics needs further attention. My findings suggest that teams containing both men and women were most likely to interact with each other in a TBL-appropriate way. Here, a TBL-appropriate way refers to students actively working together and communicating as they complete their in-class activities. The all-man team mentioned above was the only single-gender team able to do this, but only after the intervention of the instructor. Other single-gender teams fell apart or assimilated with others. The students I observed appeared to prefer to discuss non-course-related material along gender lines, which may explain the troubles that one all-man team encountered. Team 1 of Table B in Section 1 (MMM) frequently drifted off-topic with their conversations and tended to be loud and domineering at their table. Perhaps when multiple genders are represented within a group students are more likely to stay on task as to not purposefully exclude a group member with differing interests. Research has found that women are more likely to rate TBL as a positive experience than men (Wiener, Plass, & Marz, 2009). This positive experience differential may explain why the all-men teams appeared to find working as a group a challenging task.

Finally, the most consistent finding of TBL research, that students are more engaged in TBL than lecture, appears to be more complex than anticipated. Out of context, the consistent finding seems to suggest that students in TBL-based courses remain on task because they are actively engaged in classwork. However, I observed numerous ways in which students disengaged from class. Many of these forms of disengagement mirror what has been seen in lectures. Students in the TBL course were still distracted by technology. Students texted throughout class, particularly when the instructor was providing instructions or talking through problems. On a positive note, students texted less often when working in their groups. Perhaps most concerning was when students opted to remove themselves from their groups. They removed themselves mostly physically by deliberately sitting away from their team members. For one student, this did not result in him disengaging from class. He continued to work through problems and sought out assistance from the instructor when needed. Another student completely disengaged from class as well, focusing instead on playing on her tablet for the duration of class. She was keen on how to play the role of active team member when the instructor or assistants approach – she would move closer to her peers and look at their work, as if contributing. On the surface, this behavior resembles social loafing; however, she incurred no benefit from pretending to be on a team other than not being told to work with her team members. Students did not complete group assessments, so she did not loaf off of the other students in a way that enhanced her grade. A difficulty with group work in this manner is accurately discerning students who are socially loafing from those who are actually struggling with the material (Freeman & Greenacre, 2011). Without being able to analyze this student's work, I am unable to determine if this student was indeed struggling. Another potential challenge with social loafing is group size. As groups grow in size, so does the likelihood that social loafing may occur (Shimazoe & Aldrich, 2010).

Conclusion

Institutions of higher education are beginning to subscribe to the use of TBL in classes as a pedagogical strategy to get students actively engaged in their work as opposed to sitting passively and soaking up knowledge dispensed in a lecture hall. Evaluation of TBL courses has largely employed the use of surveys to measure rates of satisfaction and/or engagement. These measures have subsequently been used to support the effectiveness of the TBL approach. This study took an observational approach to study first-hand how students worked in a TBL-based course. Results suggested that previous measures of effectiveness are problematic and shadow the impact and inner workings of TBL. Additional research is needed in a variety of areas, several of which could not have been identified through survey research. Longitudinal research is also needed to investigate the impact of subsequent TBL classes on group dynamics and student engagement, especially since satisfaction has been found to be a fluid concept readily impacted by time and experiences (Reinig, Horowitz, & Whittenburg, 2011). Practical advice for instructors considering adopting a TBL approach for their courses are also offered.

References

Abdelkhalek, N., Hussein, A., Gibbs, T., & Hamdy, H. (2010). Using team-based learning to prepare medical students for future problem-based learning. *Medical Teacher, 32*, 123-129.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman.

Beatty, J. R., Haas, R. W., & Scriblimpaglia, D. (1996). Using peer evaluation to assess individual performances in group class projects. *Journal of Marketing Education, 18,* 17-27.

Carmichael, J. (2009). Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching, 38*(4), 54-61.

Chung, E., Rhee, J., Baik, Y., & A O. (2009). The effect of team-based learning in medical ethics education. *Medical Teacher, 31*, 1013-1017.

Clark, M. C., Nguyen, H. T., Bray, C. B., & Levine, R. E. (2008). Team-based learning in an undergraduate nursing course. *Journal of Nursing Education, 47*(3), 111-117.

Davis, S., & Franklin, S. (2004, August). *Assessing the impact of student learning style preferences*. Paper presented at the Physics Education Research Conference, Madison, WI.

Delucchi, M. (2006). The efficacy of collaborative learning groups in an undergraduate statistics course. *College Teaching*, *54*(2), 244–248.

Dommeyer, C. J. (1986). A comparison of the individual proposal and the team project in the marketing research course. *Journal of Marketing Education, 8,* 30-38.

Dunaway, G. A. (2005). Adaption of team learning to an introductory graduate pharmacology course. *Teaching and Learning in Medicine, 17*, 56-62.

Frame, T. R., Cailor, S. M., Gryka, R. J., Chen, A. M., Kiersma, M. E., & Sheppard, L. (2015). Student perceptions of team-based learning vs traditional lecture-based learning. *American Journal of Pharmaceutical Education, 79*(4), 1-11.

Freeman, L., & Greenacre, L. (2011). An examination of socially destructive behaviors in group work. *Journal of Marketing Education, 33*(1), 5-17.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago, IL: Aldine.

Guest, G. S., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: SAGE Publications.

Haidet, P., Morgan, R. O., O'Malley, K., Moran, B. J., & Richards, B. F. (2004). A controlled trial of active versus passive learning strategies in a large group setting. *Advances in Health Sciences Education, 9*, 15-27.

Haidet, P., Schneider, V., & Onady, G. M. (2008). Research and scholarship: Team-based learning in health professions education. In L. K. Michaelsen, D. X. Parmalee, K. K. McMahon, & R. E. Levine (Eds.), *Team-based learning for health professions education: A guide to using small groups for improving learning* (pp. 117-130). Sterling, VA: Stylus.

Hendrick, C., Giesen, M., & Coy, S. (1974). The social ecology of free seating arrangements in a small group interaction context. *Sociometry, 37*, 262-274.

Hendry, G., Heinrich, P., Lyon, P., Barrat, A., Simpson, J., Hyde, S., … Mgaieth, S. (2005). Helping students understand their learning styles: Effects on study self-efficacy, preference for group work, and group climate. *Educational Psychology, 25*(1), 395-407.

Hettler, P. L. (2015). Student demographics and the impact of team-based learning. *International Advances in Economic Research, 21*(4), 413-422.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. R. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organization Behavior and Human Decision Processes, 81*(2), 226–251.

Hrynchak, P., & Batty, H. (2012). The educational theory basis of team-based learning. *Medical Teacher, 34,* 796-801.

Johnson, D., Johnson, R., & Smith, K. (2007). The state of cooperative learning in post secondary and professional settings. *Educational Psychology Review, 19*(1), 15-29.

Karns, G. L. (2006). Learning style differences in the perceived effectiveness of learning activities. *Journal of Marketing Education, 28*(1), 56-63.

Koles, P. G., Stolfi, A., Borges, N. J., Nelson, S., & Parmelee, D. X. (2010). The impact of team-based learning on medical students' academic performance. *Academic Medicine, 85*(11), 1739-1745. doi:10.1097/ACM.0b013e3181f52bed

Krathwohl, D. R. (2009). *Methods of educational and social science research* (3rd ed.). Long Grove, IL: Waveland Press.

Lejk, M., Wyvill, M., & Farrow, S. (1999). Group learning in systems analysis and design: A comparison of the performance of streamed and mixed ability groups. *Assessment & Evaluation in Higher Education, 24*(1), 5-14.

Levine, R. E., O'Boyle, M., Haidet, P., Lynn, D. J., Stone, M. M., Wolf, D. V., et al. (2004). Transforming a clinical clerkship with team learning. *Teaching and Learning in Medicine, 16*(1), 270-275.

Lighfner, S., Bober, M. J., & Willi, C. (2007). Team-based activities to promote engaged learning. *College Teaching, 55*(1), 5–18.

MacGregor, J., Cooper, J., Smith, K., & Robinson, P. (2000). *Strategies for energizing large classes: From small groups to learning communities: New direction for teaching and learning.* San Francisco, CA: Jossey-Bass.

Mennenga, H. A. (2013). Student engagement and examination performance in a team-based learning course. *Journal of Nursing Education, 52*(8), 475-479. doi:10.3928/01484834-20130718-04

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. In L. K. Michaelsen, M. Sweet, & D. X. Parmelee (Eds.), *New Directions for Teaching and Learning, 116* (pp. 7-27). Hoboken, NJ: John Wiley & Sons. doi:10.1002/tl.330

Myers, S. A. (2012). Students' perceptions of classroom group work as a function of group member selection. *Communication Teacher, 26*(1), 50-64.

Nachmias, D., & Nachmias, C. (1987). *Research methods in the social sciences* (3rd ed.). New York, NY: St. Martin's Press.

O'Malley, K. J., Moran, B. J., Haidet, P., Seidel, C. L., Schneider, V., Morgan, R. O., et al. (2003). Validation of an observation instrument for measuring student engagement in health professions setting. *Evaluation & The Health Professions, 26*, 86-103.

Parmelee, D. X., DeStephen, D., & Borges, N. J. (2009). Medical students' attitudes about team-based learning in a pre-clinical curriculum. *Medical Education Online, 14*, 1-7. doi:10.3885/meo.2009.Res00280

Payne, B. K., Monk-Turner, E., Smith, D., & Sumter, M. (2006). Improving group work: Voices of students. *Education, 126*(3), 441-448.

Reinig, B. A., Horowitz, I., & Whittenburg, G. E. (2011). A longitudinal analysis of satisfaction with group work. *Group Decision and Negotiation, 20*(2), 215-237.

Rundle-Thiele, S., Bennett, R., & Dann, S. (2005). The successful preparation and development of future marketing professionals: A recommended methodological framework. *Journal for Advancement of Marketing Education, 7*(1), 27-35.

Searle, N. S., Haidet, P., Kelly, P. A., Schneider, V. F., Seidel, C. L., & Richards, B. F. (2003). Team learning in medical education: Initial experiences at ten institutions. *Academic Medicine, 78,* S55-S58.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(1), 63-75.

Shimazoe, J., & Aldrich, H. (2010). Group work can be gratifying: Understanding & overcoming resistance to cooperative learning. *College Teaching, 58*(2), 52-57.

Sisk, R. J. (2011). Team-based learning: Systematic research review. *Journal of Nursing Education, 50*(12), 665-669.

Su, A. Y. (2007). The impact of individual ability, favorable team member scores, and student perception of course importance on student preference of team-based learning and grading methods. *Adolescence, 42*(168), 805-826.

Thompson, B. M., Schneider, V. F., Haidet, P., Levine, R. E., McMahon, K. K., Perkowski, L. C., & Richards, B. F. (2007). Team-based learning at ten medical schools: Two years later. *Medical Education, 41*, 250-257.

Touchet, B. K., & Coon, K. A. (2005). A pilot use of team-based learning in psychiatry resident psychodynamic psychotherapy education. *Academic Psychiatry, 29*(3), 293-296.

Walters, D. E. (2012). Team-based learning applied to a medicinal chemistry course. *Medical Principles and Practice, 22*, 2-3. doi:10.1159/000342819

Wilson-Delfosse, A. L. (2012). Team-based learning: From educational theory to emotional intelligence. *Medical Teacher, 34*, 781-782. doi:10.3109/0142159X.2012.724222

Wiener, H., Plass, H., & Marz, R. (2009). Team-based learning in intensive course format for first-year medical students. *Croatian Medical Journal, 50*(1), 69-76.

Young, J. R. (2009, July 24). When computers leave classrooms, so does boredom. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com

Zgheib, N. K., Simaan, J. A., & Sabra, R. (2010). Using team-based learning to teach pharmacology to second year medical students improves student performance. *Medical Teacher, 32*, 130-135.

MINING DATA TO CREATE A SYSTEM TO IDENTIFY AT-RISK FRESHMEN

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

Abstract

Data mining is used to develop a series of three models, deployed during orientation through week six, to identify low GPA freshmen in order to improve their outcomes. Customized dashboards are developed to enable users to segment, filter, and list students to assign them to the appropriate advising plans and interventions. Previous modeling has been successful in the early identification of low GPA students and has demonstrated a strong association between learning management system (LMS) logins and GPA outcomes. Factors entered into the predictive models include advising visits, freshmen course-taking activity, LMS logins, college activity participation, SAT scores, high school GPA, demographics, and financial aid.

Introduction

The goal of this data mining effort is to predict as soon as possible, which first-time full-time freshmen students will receive a low GPA in their first term as soon as possible so they can be assigned to interventions. The fall 2012 through fall 2015 freshmen cohort students at our institution who are in the lowest first semester GPA decile had one-year retention rates that ranged from 26 to 34 percentage points lower than those in the second decile. The differences between decile 2, decile 3, and the other deciles combined were much more modest (see figure 1) The results for two-year retention were similar, with differences between decile 1 and decile 2 ranging from 24.6 to 26.3 percentage points. Again, the differences between the higher deciles were much smaller (see figure 2).

The study utilizes information gained and expands upon a fall 2015 study (Galambos 2015) that predicts fall 2015 first-time full-time freshmen GPA's by week 6 of their first semester. That study was our first to use learning management system (LMS) logins in a predictive model. It was determined that learning management system logins did, in fact, have predictive utility and were the were top GPA predictor among students having a high school GPA less than 92.0 (Galambos 2015). Further, the decision tree model provided useful early freshmen GPA estimates, as well as demonstrating differences in the set of predictors for students with different pre-college profiles, most notably high and low high school GPA. A limitation of that study was the lack of archived LMS logins, so only fall 2014 login data was available, leaving only one semester's worth of data available for modeling.

This current study combines fall 2014 and fall 2015 first-time full-time freshmen data to develop three models to predict first semester GPA and builds on methodological information gathered in the development of the previous model. (See the variable list in the appendix for a list of the measures entered into the models.) The first model uses data available on or before orientation, which includes course and major selections, to allow advisors to have an early view of students' possible GPA outcomes to aid in early advising. Course selection and early campus interactions, such as tutoring service utilization and LMS logins, were used to update the model at week three after the end of the drop and add period, and a final model was developed utilizing data through week six. K-fold cross validation was again used to avoid over-fitting, and average squared errors were used to compare the models. Based on the results of the prior study, CART and CHAID decision tree methods were used for the models with the relative importance measure used to evaluate the relative strength of the variables that are entered into the model.

Figure 1. One-year retention rates of first-time full-time freshmen by first semester GPA deciles and cohort

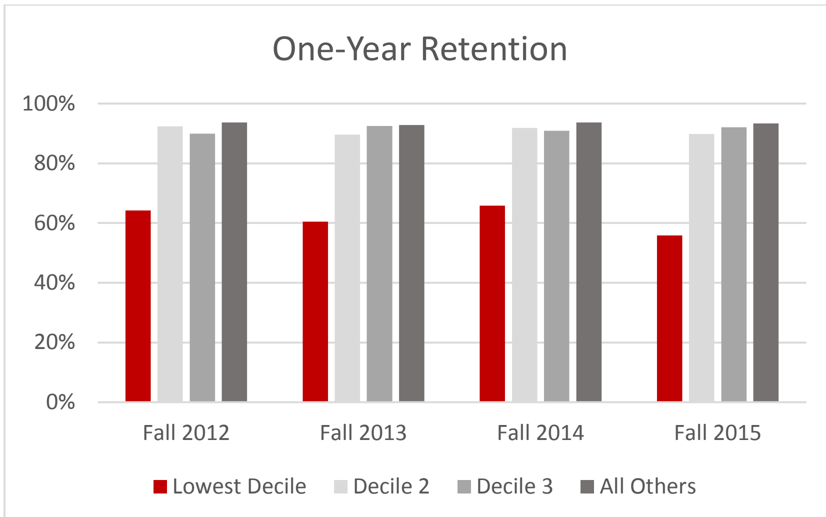
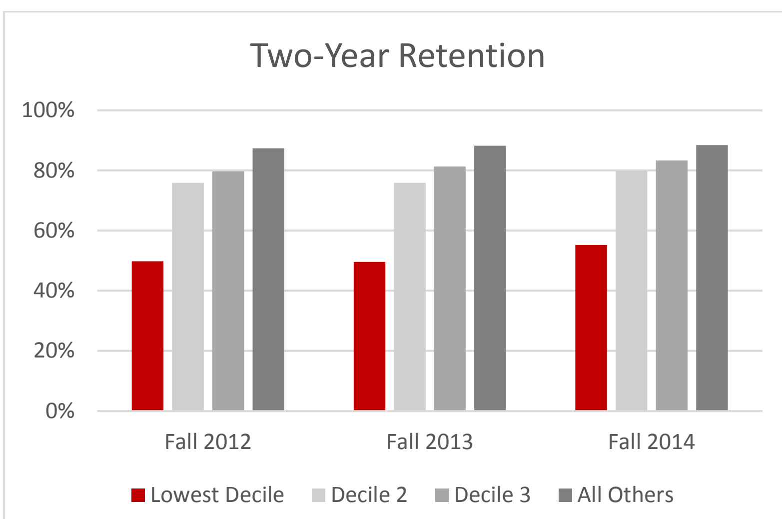


Figure 2. Two-year retention rates of first-time full-time freshmen by first semester GPA deciles and cohort



Dashboards allow users to visualize the predictions and select students for assignment to interventions. Most of the student record data are collected from the university warehouse system, however at present custodians of transaction data are contacted separately to provide LMS logins, advising and tutoring center visits, and other data. These data will eventually be placed in a designated, more easily accessible repository for both data access and archiving purposes.

Literature Review

The study has cast a wide net in terms of assembling a variety of data for use in studying academic, social, and economic factors to determine elevated risk of a low GPA, which can translate to increased risk of early attrition or longer time to degree. Consistent with the retention study of Tinto (1987), we evaluate many types of data representing students' interactions with their campus environment to determine if higher levels of campus engagement are predictive of improved freshmen outcomes. These measures of engagement include interactions with the learning management system, intramural sports and fitness class participation, and academic advising and tutoring center visits. More recently researchers at North Carolina State University presented a study demonstrating that academic achievement is improved by increasing physical activity by just one hour each week (EAB 2016).

It appears that students who are identified to be at risk in their first term and remain at the institution, continue to be at risk, with greater numbers leaving in the subsequent term (Singell and Waddell 2010). This is consistent with the results at our institution which are presented in Figures 1, 2, and 3. Methods capable of more accurate predictions will result in more effective utilization of campus resources, and higher retention and graduation rates. Course-taking behavior is also important, particularly math readiness. Herzog (2005) found math readiness to be “more important than aid in explaining freshmen dropout and transfer-out during both first and second semesters.” To account for the effects of both math readiness and course taking behavior on GPAs, we included our institution’s math placement exam results, since the placement exam is administered to all newly enrolled students at our institution. Additionally, we tallied the number of credits of high failure rate courses in which the students were enrolled. Herzog also focused on both merit and need-based aid, and the role that the interaction of aid and academic preparedness plays in student retention. Living within a 60 mile radius of the institution, the percent of students at a high school who take the SAT, along with the percentage at the high school receiving free lunches was explored by Johnson (2008) underlining the need to examine the role of the secondary school and socio-economic factors in developing a model. Persistence increases among students closer to the institution and not surprisingly, decreases among those who were from schools having a high percentage of students receiving free school lunches. The role of differing stop-out patterns exhibited by grant, work-study, and loan recipients (Johnson 2010) demonstrated that grants have the highest positive effect on persistence, but its effect decreases more than that of loans after controlling for other factors.

Resource utilization was studied (Robbins et al. 2009) using a tracking system. Services and resources were grouped into academic services, recreational resources, social measures and advising sessions, with all but social measures demonstrating positive associations with GPA even after controlling for other demographic and risk factors. We have included tutoring center and academic advising visits, and, as previously mentioned, the recreation center usage. The relationship of learning management system usage with student outcomes is of particular interest. A study of five online biology courses (Macfadyen and Dawson, 2010) examined a variety of LMS tracking measures including the number of discussion messages, new discussion posts, assignments read, and time spent on assignments. Of the 22 LMS metrics evaluated, 13 were significantly correlated with the students' final grades. Further, a regression analysis found that total discussion posts, total mail messages sent, and total assessments completed accounted for 33% of the variation in student achievement scores in the course, and logistic regression correctly categorized as "at risk" 17 of 21 (80.0%) of students who ended up failing the course. In 2007 Romero, et. al. examined a number of data mining methods to demonstrate how they can be used to study outcomes in an open-source LMS online course environment.

These papers have demonstrated that researchers are examining a range of factors in studying and modeling risk. The research highlights the fact that student success is a complex interaction of student engagement, academic service utilization, financial metrics, demographics, combined with student academic characteristics that include high school GPA and SAT scores. Data mining is ideal for developing a model with a large diverse number of predictors.

Methodology

A broad list of data was selected for model development. The more traditional data include demographics, pre-college characteristics, and financial aid measures. In addition to those items the list of college measures includes major groupings, number of AP courses accepted for credit, and number of courses with large proportions of D, F, and W grades, i.e., high DFW courses. A course was coded as a high DFW course if it has an enrollment of at least 70 students with 10% of its grades consisting of D's, F's, or W's. Service utilization data includes Learning Management System (LMS) logins, tutoring center visits, academic advising interactions, and recreation center usage. Studying the use of LMS logins is consistent with research that has shown that engagement with the campus environment improves student outcomes. LMS logins were tabulated as follows. One login per course per hour per student was counted, so each student can have a maximum of 24 logins in each course per day. This eliminated multiple logins in the data that occurred just seconds apart. Total logins (using the previous definition) were tabulated for each time period, weeks 1 to 3, and weeks 1 to 6. In addition, total logins were divided by the number of courses utilizing the Learning Management System in which the student was enrolled to create an additional “logins per course” metric. The optimal method for utilizing the LMS data remains an area of active research. Other measures include the average SAT scores of the high schools to control for high school GPA, a variety of financial aid measures, number of enrolled credits grouped by STEM and non-STEM, and AP courses accepted for credit. (See the Appendix for a more complete listing of the data.)

Considerable effort was expended in developing the model to predict the fall 2015 freshmen GPA at week 6 (Galambos 2016). Five different methods were compared with gradient boosting, classification and regression trees (CART), and chi-squared automatic interaction detection (CHAID) having the lowest average squared errors in that order. Because the gradient boosting method yields scoring code, with no explicit, easily understood algorithm or decision tree, and additionally did not demonstrate a substantive error rate reduction, it was not used. Being able to understand how the predictors contribute to student GPA outcomes is useful for selecting and assigning students to interventions and monitoring measures to help keep students on track. The graphic decision tree display is compelling in that regard.

With LMS data available for both fall 2014 and fall 2015, two years of data were used to develop the three models to predict the fall 2016 freshmen GPA’s. The total number of first-time full-time fall 2014 and fall 2015 freshmen was 5,664 after 34 students who withdrew prior to the end of the term were removed from the sample. In order to avoid overfitting the model the data are typical divided into training and validation sets. The model is developed using the training set after which the model is run on the hold out validation sample. We expect similar error results in both the training and validation sets if the model is performing well. Our sample has close to 5,700 students, which may seem sufficient for a 60/40 training to validation data split, however if one considers that over 50 variables are being entered into the model and we are mainly focused on obtaining accurate predictions for the bottom GPA decile, only about 350 students in the group of interest would be left in our sample. As with the previous year's model, K-fold cross validation was used, which allows us to subdivide the sample into 5 groups or folds and run the model five times using 80% of the data and then validating it on the remaining 20%, with a different hold out sample used each time the model is run. Figure 3 shows the 5-fold cross validation scheme. The error results are obtained by taking the average of the five average squared errors (ASE)[1] generated for the training and validation samples for each fold.

Figure 3. K-fold cross-validation sampling design.[2]

K=1	Train	Train	Train	Train	Validate
K=2	Train	Train	Train	Validate	Train
K=3	Train	Train	Validate	Train	Train
K=4	Train	Validate	Train	Train	Train
K=5	Validate	Train	Train	Train	Train

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

[2] From Galambos N., (2015). Using data mining to predict freshmen outcomes. *42nd NEAIR Annual Conference Proceedings*, February 2016, p. 89.

The current data were modeled using both CART and CHAID[3]. CART does an exhaustive search for the best binary split at each node. For interval targets the variance is used to assess the splits; for nominal targets the Gini impurity measure is used. The result is a set of nested binary decision rules to predict the outcome. CHAID on the other hand uses the chi-square test to determine categorical splits and F tests for intervals. It allows multiple splits in continuous variables and allows categorical data to be split into more than two categories.

Results

In the fall 2015 study the focus was on identifying measures that can be used to predict freshmen GPA by mid-semester and how those predictors differed by student academic profiles or characteristics. For this set of models, the focus is on the small sections of the decision trees providing the low GPA predictions. The resulting algorithms will be used to assign GPA predictions to the fall 2016 freshmen cohort data for use by the appropriate stakeholders. The average squared errors for the cross validation results, presented in Table 1, are similar to those obtained for the fall 2015 model, but with slightly more concordance between the training and validation errors. A GPA prediction was made on day 1 and forwarded to advisors and others in contact with students so they could take early action or monitor students' progress. Though the average ASE of the CHAID model for day 1 was slightly higher than that of the CART model, the decision was made to use the CHAID method for the day 1 low GPA model, since it had a high level of agreement between training and validation results for the nodes of interest. One of

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

the predictors, the number of high DFW rate courses in which a student is enrolled, is highly actionable at the beginning of the term.

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

	Day 1 Model (CHAID)		**Week 3 Model (CART)**		**Week 6 Model (CART)**	
K Folds	**Validation ASE**	**Training ASE**	**Validation ASE**	**Training ASE**	**Validation ASE**	**Training ASE**
1	0.46	0.41	0.44	0.46	0.43	0.46
2	0.48	0.41	0.45	0.44	0.43	0.44
3	0.50	0.40	0.45	0.46	0.44	0.43
4	0.51	0.40	0.45	0.43	0.46	0.43
5	0.56	0.39	0.51	0.43	0.51	0.42
Average ASE	**0.50**	**0.40**	**0.46**	**0.44**	**0.45**	**0.44**

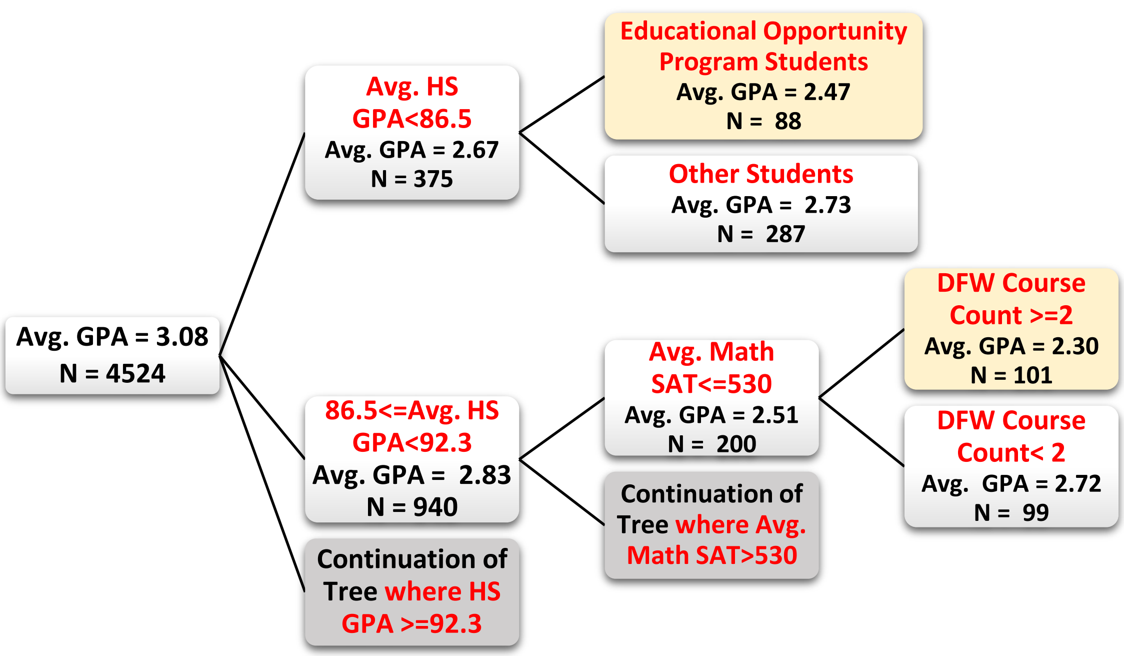


Figure 4. Low GPA Portion of CHAID Model for the Predicting Freshmen GPA on Day 1.

For each of the three models the low GPA section of the corresponding decision tree is shown. Nodes having an average GPA of less than 1.00 are highlighted in orange and those with an average GPA of 1.00 to below 2.50 are highlighted in yellow. Lists of fall 2016 students selected by the decision rules of the highlighted nodes were provided to the appropriate entities on campus. The gray nodes indicate where sections of the decision trees have been truncated to facilitate the graphic presentation, and because they do not contain any nodes in the low GPA range of interest. The node frequencies reflect the fall 2014 and fall 2015 training samples which were used for the model. The model predictors are displayed in red in each node, below which is the predicted GPA for students falling within the corresponding decision rule for the node.

The Educational Opportunity Program (EOP) which figures prominently in the day 1 model (Figure 4) is a program for students whose circumstances, both economic and educational, may have limited their options for obtaining a post-secondary education. Students accepted into the program are typically from historically disadvantage backgrounds and have demonstrated potential for finishing college although they may not have been accepted through the traditional admission process. The program provides financial assistance, tutoring, and mentoring. Because students are admitted to that group by virtue of their lower academic profile, it is not surprising that in the day 1 model some of the EOP students, those with a high school GPA below 86.0, are predicted to have low GPA outcomes. Students having a high school GPA in the 86.5 to 92.3 range, math SAT scores of 530 or less, and are enrolled in 2 or more high DFW rate courses are also predicted to have a low GPA. The average GPA prediction for the EOP students with the lowest high school GPA is 2.47, and is 2.30 for the students with a slightly higher high school GPA, low math SAT scores and two or more high DFW courses. Those are the lowest average

GPA nodes in the entire model. There are none as low or lower within the nodes not displayed in figure 4. A list of those students was provided to the campus stakeholders who provided them with tutoring and peer mentors, as appropriate.

The score distribution table, table 2, part of the decision tree output, has 20 equally spaced bins created by dividing the interval between the highest and lowest predictions by 20, and presents the average GPA and number of students in each interval. Bins with no observations are removed from the table. The model scores are calculated by taking the mid-point of each interval. Since the table shows the number of students at each average GPA level, it can assist in choosing GPA cut points for intervention groups. The number of values in each row are based on a fall 2014 and fall 2015 training sample used for the model.

Table 2. Day 1 Model Score Distribution Table

Prediction Range	Average GPA	N	Model Score
3.80 - 4.00	3.94	3	3.90
3.60 - 3.80	3.71	233	3.70
3.40 - 3.60	3.56	426	3.50
3.20 - 3.40	3.32	1312	3.30
3.00 - 3.20	3.04	932	3.10
2.80 - 3.00	2.91	362	2.90
2.60 - 2.80	2.74	981	2.70
2.40 - 2.60	2.46	147	2.50
2.20 - 2.40	2.31	104	2.30
2.00 - 2.20	2.18	8	2.10
1.80 - 2.00	1.96	5	1.90
1.60 - 1.80	1.61	1	1.70
1.40 - 1.60	1.52	2	1.50
0.60 - 0.80	0.73	4	0.70
0.40 - 0.60	0.50	2	0.50
0.20 - 0.40	0.34	5	0.30
0.00 - 0.20	0.00	1	0.10

As part of the modeling process relative importance measures are calculated and provided as part of the output. “The relative importance measure is evaluated by using the reduction in the sum of squares that results when a node is split, summing over all of the nodes.[4] In the variable importance calculation when variables are highly correlated they will both receive credit for the sum of squares reduction, hence the relative importance of highly correlated variables will be about the same. For that reason, some measures may rank high on the variable importance list, but do not appear as a predictor in the decision tree."[5] The top importance measures have been included in tables presented below and include measures that may only appear in the portions of the decision trees that have nodes with higher average GPA's. For the day 1 model, high school GPA heads the list, followed by average high school SAT scores, which controls for high school quality, SAT math plus critical reading, size of scholarship received, math placement scores, total DFW STEM credits, and overall total STEM credits.

Table 3. Variable Importance Table for Day 1 Model

Variable	Relative Importance
High School GPA	1.0000
Avg. High School SAT Critical Reading, Math Score	0.5208
Avg. High School SAT Score	0.4921
SAT Math and Critical Reading Score	0.2939
Total Disbursed Scholarship Funds	0.2834
Math Placement Score	0.2806
Total DFW STEM Credits	0.2557
Total STEM Enrolled Credits	0.2145

At week 3 the second model was developed. High school GPA again was the measure that was most associated with average GPA outcomes, with total LMS logins at week 3 associated

[4] SAS Institute Inc. 2014. *SAS® Enterprise Miner™ 13.2: Reference Help*. Cary, NC: SAS Institute Inc. p. 794.
[5] Galambos N., (2015). Using data mining to predict freshmen outcomes. *42nd NEAIR Annual Conference Proceedings,* February 2016, p. 89.

with the GPA outcomes for students with a high school GPA less than 94 (figure 5). Those with less than 61 logins through week3, who attended a school where the average combined SAT math, critical reading, and writing score was less than 1570, and finally had less than 2.1 logins per course as of week three, had an average GPA of slightly below 1.00. If instead they had more than 2.1 logins per course at the week 3 time point, and 5 or more credits in high DFW rate courses, they were predicted to have a GPA of 2.27. If they went to a high school that had an average math, critical reading, and writing exam over 1570 and less than 5.2 LMS logins per course in the first 3 weeks, their average GPA was 2.30.

Table 4. Score Distribution Table for Week 3 Model

Prediction Range	Average GPA	N	Model Score
3.48 - 3.61	3.59	525	3.55
3.35 - 3.48	3.46	383	3.41
3.22 - 3.35	3.29	705	3.28
3.08 - 3.22	3.16	514	3.15
2.95 - 3.08	2.98	1006	3.02
2.82 - 2.95	2.90	694	2.88
2.68 - 2.82	2.72	290	2.75
2.55 - 2.68	2.66	120	2.62
2.28 - 2.42	2.35	93	2.35
2.15 - 2.28	2.27	184	2.22
0.95 - 1.09	0.95	10	1.02

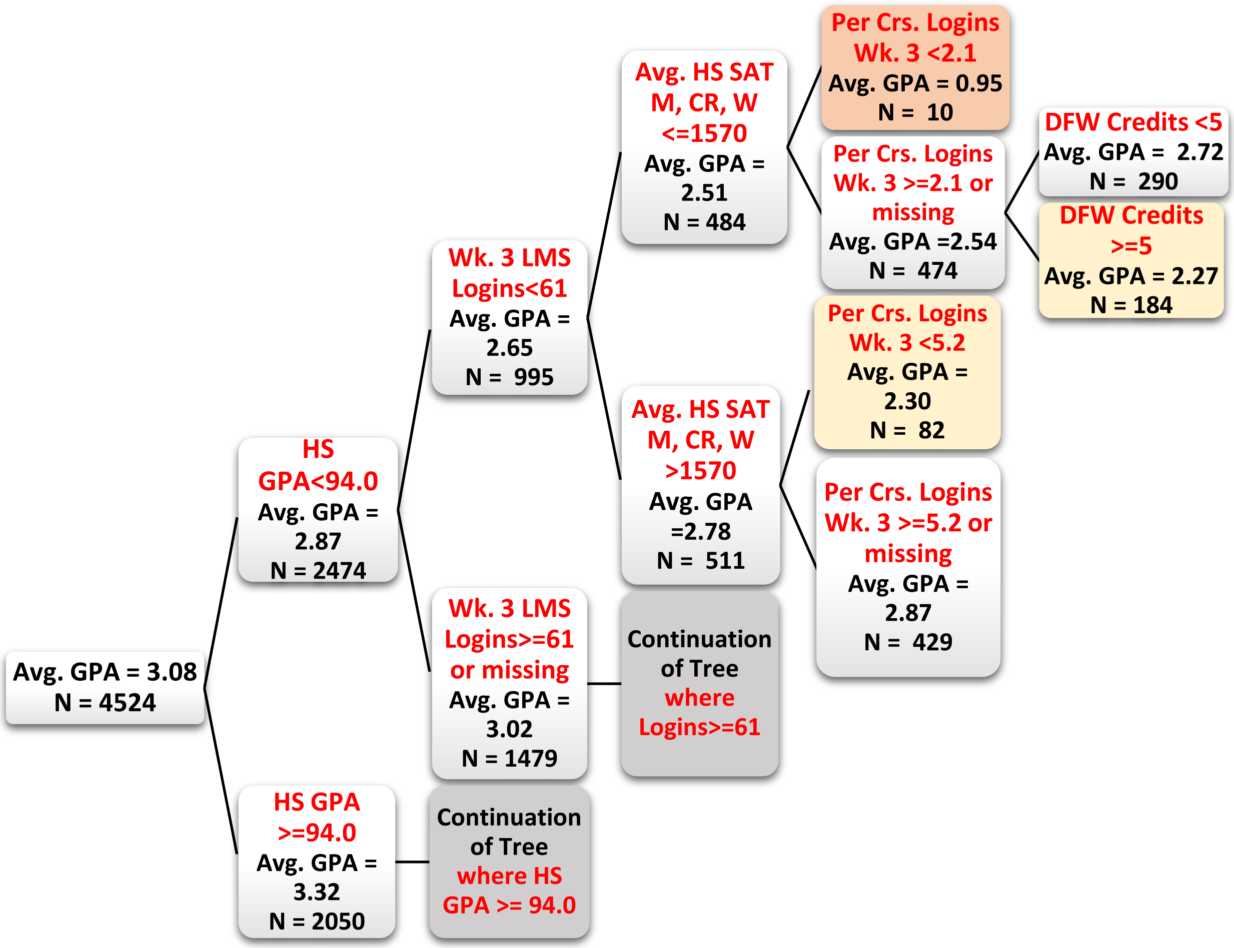


Figure 5. Low GPA Portion of the CART Model Predicting Freshmen GPA: Week 3.

The high school GPA was again at the top of the relative importance list (table 5) and LMS logins also appear on the list with high variable importance scores. The inclusion at week 3 of actual data on the students' interactions with the campus and their courses has strengthened the magnitude of the importance measures and has altered the variables included on the list considerably. Note that although ethnicity, geographic location, and tuition did not appear on the day 1 model, those items (and other demographic and pre-college measures) were, in fact, entered into the day 1 model.

Table 5. Variable Importance for Week 3 Model

Variable	Relative Importance
High School GPA	1.0000
IPEDS Ethnicity	0.8600
Academic Level	0.8281
Area of Residence at Admissions--6 Categories	0.8152
Total LMS Logins at Week 3	0.8016
Major Type—Major, Undeclared, Area of Interest	0.7516
Residency Tuition	0.7483
SAT Math and Verbal Combined	0.6246
Per Course STEM LMS Logins, Week 3	0.5867
Per Course STEM Total LMS Logins, Week 3	0.5836
Total DFW STEM Units	0.4527
Avg. SAT CR+M+W Avg. for the High School	0.4417

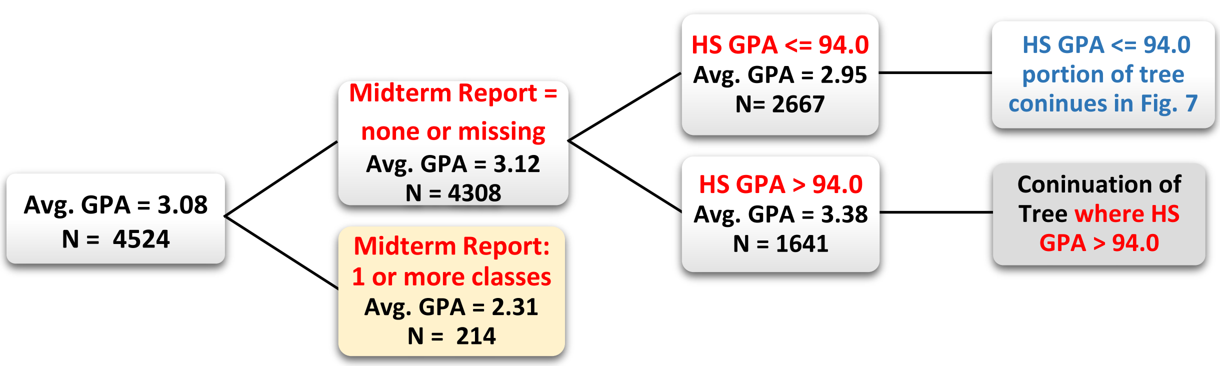


Figure 6. Low GPA Portion of the CART Model Predicting Freshmen GPA at the End of Week 6: Part 1.

The final model, using data as of the end of week 6, is presented in two parts, shown in figures 6 and 7. Part 2, figure 7, continues from top, right node with blue text.

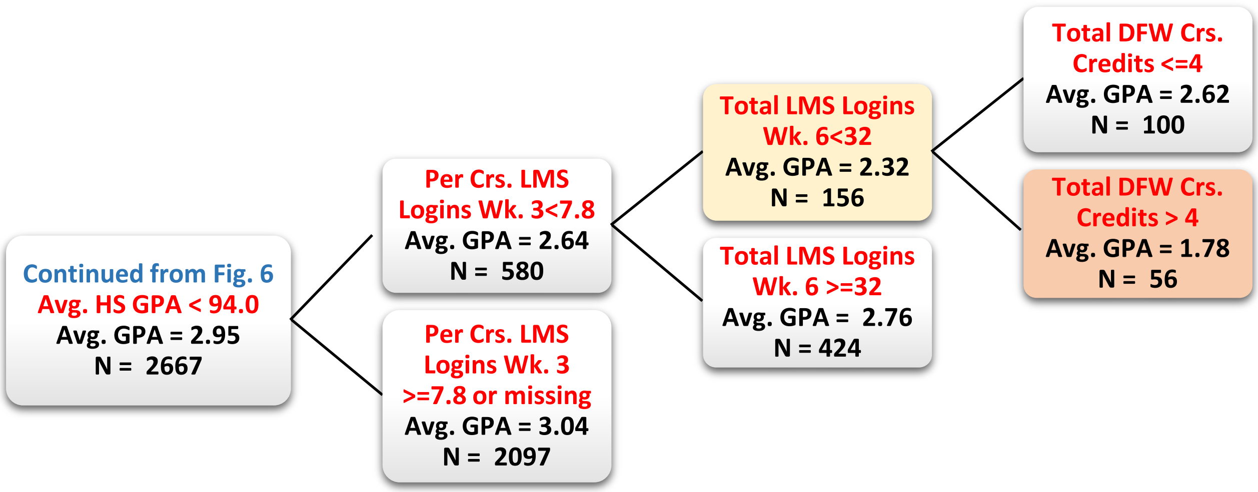


Figure 7. Low GPA Portion of the CART Model Predicting Freshmen GPA at the End of Week 6: Part 2.

In the week 6 model the top measure associated with the GPA outcome was the midterm grade report. The midterm report is a requested from professors by Academic Advising around week 6. Not all professors respond, but those that do provide midterm grade information for students in their classes. Academic Advising reaches out by email to all students on the list. The midterm report data element in the model represents the number of courses for which academic advising received a report pertaining to the student. The fall 2015 midterm report list was the first one available for the modeling process, hence the measure is missing for fall 2014. The average GPA for the midterm report node is 2.31. Those for whom there was no midterm report and additionally had a high school GPA of 94.0 or less, the model continues in figure 7. For those low high school GPA students who additionally had low LMS logins at week 3 and week 6, and more than 4 credits of high DFW rate courses, the predicted GPA was 1.78 (see figure 7).

In the week 6 model, an LMS login measure has risen to the top of the variable importance table and we see the table populated with a number of LMS login measures along with high school GPA and the midterm report.

Table 6. Week 6 Model Score Distribution Table

Prediction Range	Average GPA	N	Model Score
3.406 - 3.556	3.50	1091	3.48
3.257 - 3.406	3.38	295	3.33
3.107 - 3.257	3.21	302	3.18
2.958 - 3.107	2.99	2040	3.03
2.808 - 2.958	2.96	157	2.88
2.509 - 2.659	2.64	367	2.58
2.360 - 2.509	2.45	198	2.43
1.762 - 1.912	1.78	56	1.84
0.567 - 0.716	0.57	16	0.64

Although there is some variety in the measures predicting the GPA outcomes, we find high school GPA, number of high DFW rate courses, and LMS logins playing a prominent predictive role in all three models. In terms of the lowest high school GPA EOP students, we notice that the EOP student group did not appear again in the week 3 and week 6 models. As previously discussed, students in that program receive tutoring, peer mentoring, and other academic assistance, so clearly once that the semester progressed those EOP students as a group were no longer predicted to have a low GPA. In fact, the academic support program for the EOP students can serve as a model in designing interventions for other students. With profiles that resulted in some of them being predicted to have a low GPA in the day 1 model, it is important to note that the fall 2009 freshmen cohort EOP students had a six-year graduation rate of 79.7%, well surpassing 68.3 %, the rate for the entire fall 2009 cohort.

Table 7. Variable Importance for Week 6 Model

Variable	Relative Importance
Per Course Total LMS Logins, Week 3	1.0000
High School GPA	0.9731
Total LMS Logins, Week 3	0.8932
Total LMS Logins, Week 6	0.7606
Academic Level	0.6842
Midterm Report	0.6300
Area of Residence at Admissions--6 Categories	0.6102
Dorm Housing Indicator	0.5923
Women in Science and Eng. Program	0.5848
Total LMS Non-STEM Logins, Week 6	0.5047
Per Course Non-STEM LMS Logins, Week 6	0.4952
Non-STEM Total Logins, Week 3	0.4345
Per Course Total LMS STEM Logins, Week 3	0.4339

Data Delivery

Samples of data delivery methods with filters and the ability to drill down to the student level data can be found in the Appendix. Dashboards can easily be customized depending upon the user. Advisors may want to be able to easily find the students in various predicted low GPA groups, then drill down and view their schedules and other information. As evidenced by the graphs in the introduction, intervening early in imperative because roughly half may be gone by the end of their second year. Departments may also want to determine how many majors they have who are predicted to have low GPA's to motivate their own interventions and department advising. Since the number of students predicted to be on the lowest end of the GPA spectrum is only 10 to 15 percent of the freshmen, providing the data in spreadsheet form can also suffice.

Conclusion

The modeling process has demonstrated that measures most strongly associated with low GPA outcomes are related to how individuals perform as students, as evidenced by the variable importance scores of the high school GPA and LMS logins. Providing predictive information to those providing support services can head off a potentially damaging low GPA outcome. Additionally, being alerted to the lower high school GPA students, who may be taking multiple difficult courses may help advisors to be more pro-active in terms of helping students with challenging course schedules. Peer mentoring and tutoring have already been suggested to some such students at our institution.

Since the data used is being pulled in from many sources, the next logical step is to create a repository allowing easier access, which will in turn streamline the modeling process. Additionally, the data being collected contains information on services being provided to students such as tutoring and advising. These data not only have predictive utility and can be used to track interventions, but are also a gold mine of information that can be used to understand and study what students are utilizing the services we are providing to enable them to succeed.

References

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

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

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

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

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

"Gym time correlated with graduating, making better grades." Retrieved from:

https://www.eab.com/daily-briefing/2016/05/05/gym-time-correlated-with-graduating-making-better-grades?WT.mc_id=Email|Daily+Briefing+Headline|DBA|DB|May-05-2016|||||&elq_cid=1621083&x_id=003C000001xsbjvIAA . n.p. May 5, 2016.

Galambos N. (2015). Using data mining to predict freshmen outcomes. *42nd NEAIR Annual Conference*

Proceedings, February 2016, p. 89.

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

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

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

Macfadyen LP, Dawson S. Mining LMS Data to Develop an "Early Wrining System" for Educators: A

Proof of Concept. *Computers & Education*. 2010, 54: 588-599.

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

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

Romero C, Ventura S, Garcia E. Data Mining in Course Management Systems: Moodle Case Study and

Tutorial. *Computers & Education*. 2008, 51: 368-384.

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

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

Stater M. (2009). The impact of financial aid on college GPA at three flagship public institutions.

American Educational Research Journal. 46:782-815.

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

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

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

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

Appendix

Variable List

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

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

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

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

Financial Aid Measures

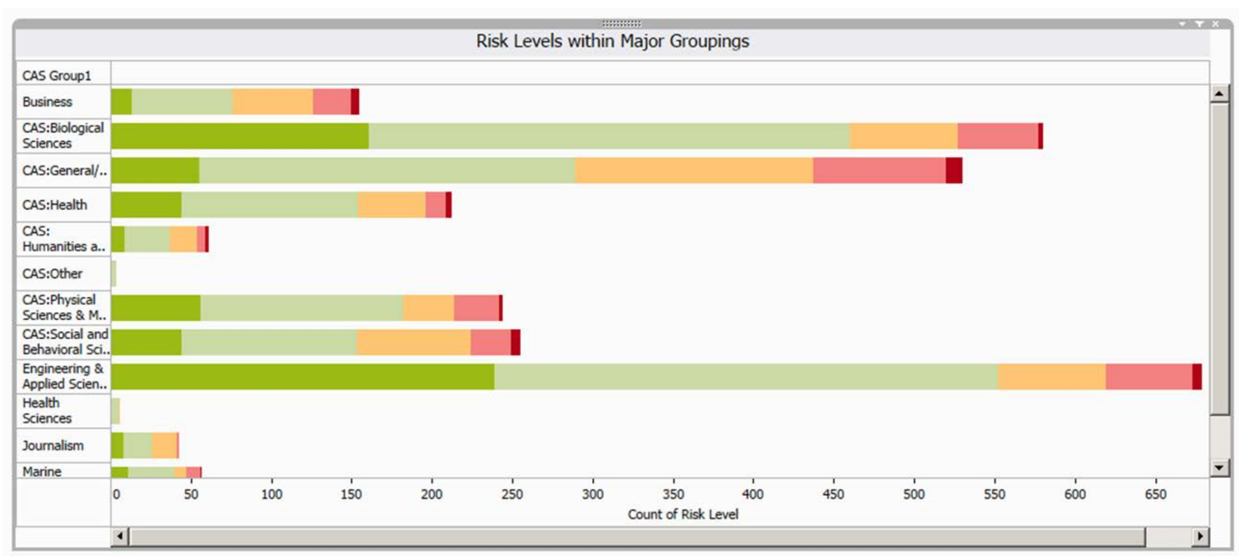
Aid disbursed in the Fall 2014 and Fall 2015 academic years
Total grant funds received
Total Loans recorded by the Financial Aid Office
Total scholarship funds received
Total work study funds received
Total athletics aid received
Athletic aid, grant, loan, PLIS loan, subsidized/unsubsidized loan, scholarship,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

Services/Learning Management System (LMS)

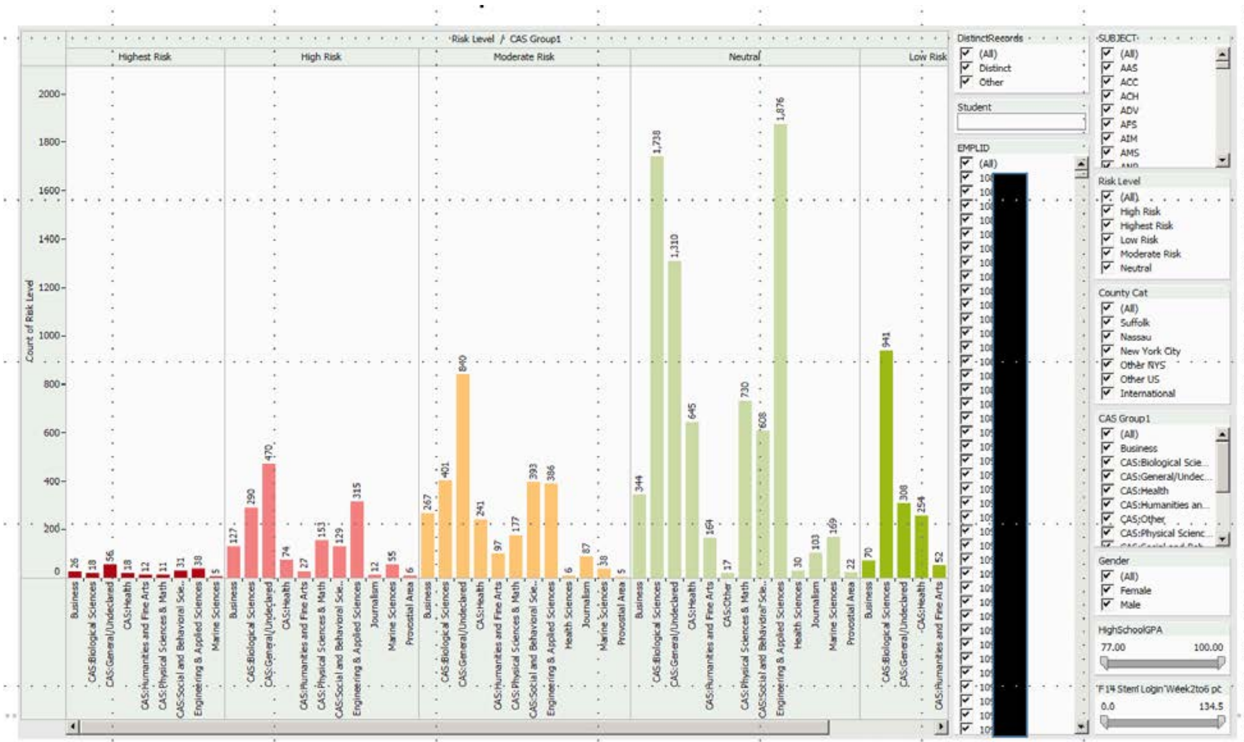
Advising Visits/Tutoring Center Usage
Tutoring center appointment no shows
Number of STEM Course Center Visits, weeks 1 to 6
Number of non-STEM Course tutoring Center visits, weeks 1 to 6
Advising Visits during week 1- 3
Advising visits during weeks 3 – 6
Course Management System Logins
F14 and F15 Stem Logins
F14 and F15 NonStem Logins Weeks 1 -3
Non-STEM course related logins during weeks 3 - 6
Non-STEM Course related logins during week 1 -3
STEM Course related logins during week 1 -3
STEM Course related logins during weeks 3 to 6
Number of STEM course logins per STEM course using the CMS, weeks 1 – 6.
Number of non-STEM course logins per non-STEM courses using the CMS, weeks 1 – 6.

Dashboard Samples

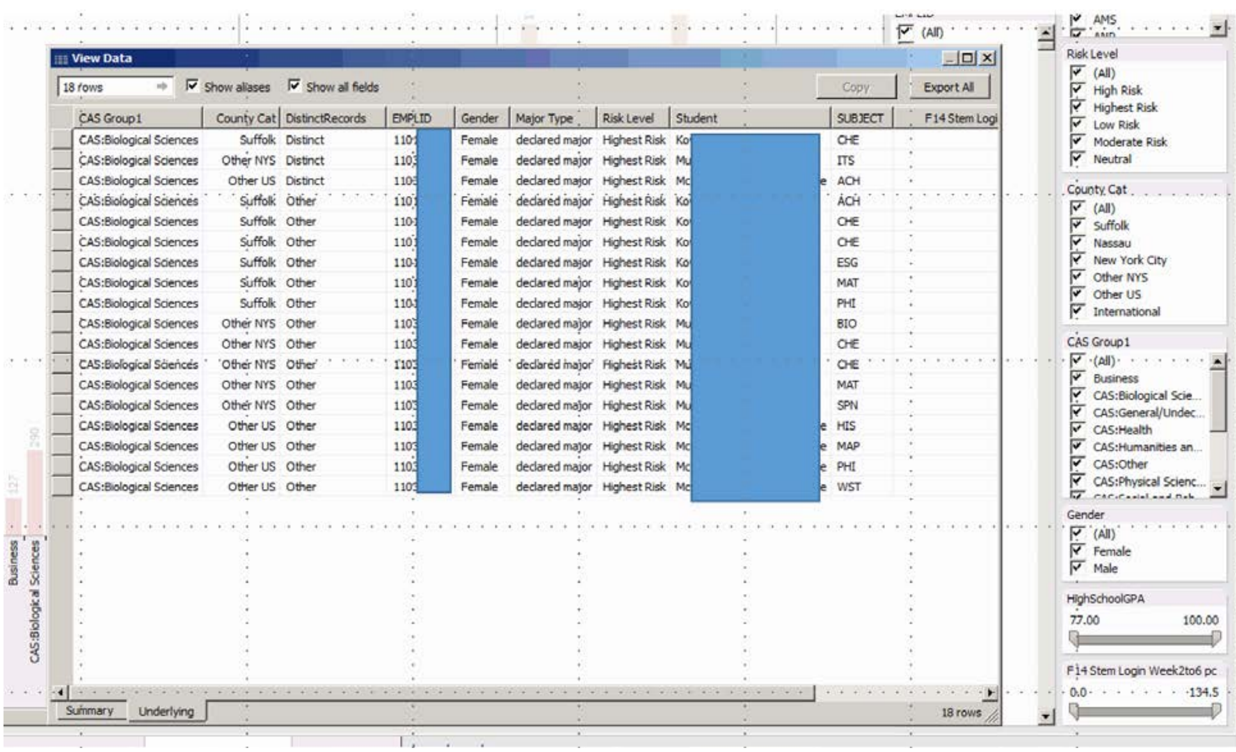
Red and pink colors represent the lowest GPA levels. Checkbox lists allow filtering.



Users can choose majors, id number, can find a student by entering their name. Sliders at the bottom right allow selection of a GPA and/or LMS login range.



Double clicking on any bar above, allows drilling. down to student data.



DEVELOPING A COMPREHENSIVE MEASURE OF FACULTY ACTIVITY

Jennifer L. Snyder

Graduate Research Assistant

Office of Institutional Research and Effectiveness

University of Delaware

Abstract

With the rising costs of higher education, the public, legislatures, and higher education administrators have shown a growing interest in faculty activity. The current analysis is from Stage 1 of a larger project designed to develop a national study of discipline-level faculty activity. Open-ended interviews with chairs and directors at public universities revealed a need for a comprehensive analytical tool that would help them better understand their unit's activity. However, the perceived usefulness of such a study also comes with certain apprehensions that must be considered when designing the theoretical and methodological framework for the study.

In 2012, David C. Levy, president of the Education Group at Cambridge Information Group and former Chancellor of The New School, published a critical opinion piece in *The Washington Post* asking, "Do college professors work hard enough?" (Levy, 2012). While he generally supports the designated workload of faculty at research-oriented institutions, Levy criticizes the culture of higher education that allows a faculty member at a teaching-oriented institution to receive a salary of over $80,000 "based on a workload of 15 hours of teaching for 30 weeks" (Levy, 2012). It only takes a brief scan of the online comment section attached to the article to understand the readers' frustration with Levy's suggestions. While research has consistently shown that college professors clock in well over fifty hours per week on average (Jacobs, 2004), the conversation regarding the rising costs of higher education often turns to faculty salaries and faculty activity. When Vice President Joe Biden visited a high school in Doylestown, Pennsylvania, he responded to the question about what is driving up the cost of higher education with, "[One reason is that] salaries for college professors have escalated significantly" (Metra4.com, 2012). During a time of rising costs, various interested parties (higher education administrators, legislatures, tax payers, students and their parents, etc.) are all, understandably, concerned with how money is being spent. Faculty workload and, more specifically, faculty productivity, is not immune to this concern. However, questions about faculty activity have long been an important component to the conversations surrounding higher education. In 2001, Robin Wilson published the article "It's 10a.m. Do you Know Where Your Professors Are?" in *The Chronicle of Higher Education*. The article was written in response to new regulations at Boston University that required faculty to work from their campus offices four days a week—a significant change for many faculty, who often spend a considerable amount of time working in their home offices or conducting research in libraries or archives (Wilson, 2001). The general argument from BU administrators was that if faculty were being paid to work at the institution, then they should be held more accountable for the work that they are doing.

While these conversations are separated by more than a decade, they embody a familiar narrative: a call for more transparency and accountability in higher education. Specifically, there is growing interest for a better understanding of how faculty are spending their time outside of the classroom. One of the first steps in deepening this understanding is to educate legislatures, taxpayers, students, and their parents that faculty activity goes well beyond the teaching that occurs in the classroom. The present study aims to expand this conversation. In a time of a growing reliance on data to make informed decisions in higher education, administrators need to ensure that the tools used to collect this data take into account the nuance of what is being studied. Most importantly, the methods of data collection must consider both the goals of the end-users and the concerns of those being studied.

Measuring Faculty Activity

In 1992, the Office of Institutional Research and Effectiveness (then the Office of Institutional Research and Planning) at the University of Delaware developed a national benchmarking study of instructional costs (Middaugh, 2001). The study sought to answer the question "Who is teaching what to whom and at what cost?" (Middaugh, 2001). While the study focuses on faculty activity within the classroom, it is unable to capture the nuanced nature of faculty activity, specifically that which is occurring outside of the classroom.

In 2001, under a grant from the Fund for Improvement of Postsecondary Education (FIPSE), the office was commissioned to expand the study to measure faculty activity outside of the classroom (Isaacs and Middaugh, 2004). With the help of an advisory committee, the office developed a survey to capture selected measures of faculty outputs, or products, in the areas of teaching, scholarship, and service. In spring 2003, the office conducted the first full data collection cycle of the Out-of-Classroom Faculty Activity Study (FOCS Study) (Isaacs and Middaugh, 2004).

For many years the Out-of-Classroom Faculty Activity Study offered useful benchmarking data to participating institutions; however, in 2008, the study was placed on an indefinite hiatus. In 2015, the Office of Institutional Research and Effectiveness's Higher Education Consortia (HEC) expressed a renewed interest in developing a study of faculty activity. This new study, the Faculty Activity Trifecta (FACT) Study, is being designed to reimagine the way that faculty activity outside of the classroom is measured. The “trifecta,” in addition to emphasizing a holistic view of faculty activity as teaching, scholarship, and service, considers the idea that faculty activity is more than just an exhaustive list of outputs. Instead, faculty activity outside of the classroom, especially when measured at the discipline level, is extremely nuanced, and ultimately, is an essential component to the achievement of an institution’s mission and goals.

A Culture of Accountability

Regulating faculty work has always been a major policy issue in higher education (Porter and Umbach, 2001). Research over the last few decades (along with recent media attention) indicates a continual call for transparency and accountability, specifically in regards to how faculty members are spending their time. This current culture is evident of sentiments left over from the 1990s in which state legislatures established required minimums on instructional workload (Terpstra and Honoree, 2009; Jacobs, 2004). In the early 1990s, many states adopted stricter higher education accountability laws (AAUP, 1996). Additionally, as higher education started moving towards a more managerial approach to learning, it became necessary to adhere to more a market-based model of consumer satisfaction or output (AAUP, 1996). Thus, with the rising costs of higher education, the public (i.e. taxpayers) wants to see that public institutions are making every effort possible to cut costs (Miller, 1994). Often, when it comes to cutting costs, the public and legislatures tend to offer “common-sense” solutions: either increase faculty productivity or decrease state funding (Miller, 1994).

While research has consistently shown that full-time faculty do have lengthy workweeks (e.g. Jacobs, 2004; Miller, 1994), *how* they are spending that time has long been contested.

Specifically, the flexibility of faculty schedules means that faculty members are often working outside of their offices and during-non traditional work hours, so they rarely have a clear division between leisure and work. This can lead to many people (both inside and outside of academia) being skeptical of the faculty workweek (Yuker, 1984).

While recent legislation has called for an increase in individual instructional workloads, those within higher education understand that the achievement of an institution's mission and goals depends on a trifecta of faculty activity: instruction, scholarship, and service. Consequently, a need has arisen for institutional researchers to develop a comprehensive measure of faculty activity. While many measures of faculty activity currently exist, they fall short in providing a complete picture of how instruction, scholarship, and service form a symbiotic trifecta of faculty activity. Ultimately, the proposed Faculty Activity Trifecta Study would be a supplemental study to the Delaware Cost Study in that it would provide a discipline-level comparative analysis of faculty activity outside of the classroom. This study aims to further understand how the FACT Study fits in to the current conversation in higher education regarding the measurement and analysis of faculty activity, specifically the activity conducted outside of the classroom.

Methods

The empirical data for this study were collected as part of Stage 1 of the FACT Study. While Stage 1 is a larger study designed to develop a comprehensive measure of faculty activity, the current analysis focuses on the underlying assumptions of developing such a measure. This investigation was designed to examine the perceived usefulness of the FACT Study for unit heads (department chairs and school directors) at institutions of higher education.

The decision to study unit heads was made intentionally with regard to the unique benefits that this population offers. Unit heads occupy a very distinct position in higher education. By functioning in a duel role as both a faculty member and as an administrator, they act as a gateway between the university's formal administration and the faculty members within their unit. Additionally, when developing a measure of faculty activity, unit heads have direct experience both with being evaluated and with evaluating their faculty's activity. Methodologically, it is beneficial to include future participants in the design phase when developing a new metric (Yuker, 1984). This is especially true for faculty administrators because they have an insider's perspective on the complex nature of faculty activity. Because of the sometimes contentious nature of measuring faculty activity, it is important to have the support of unit heads moving forward. By including faculty in the construction of the FACT Study, future participants can be assured that the opinions and suggestions of those who will be studied were taken into account when the study was designed.

Stage 1 focused on unit heads from a randomly selected sample of sixteen public colleges and universities in the United States. The sixteen institutions were sampled from the four primary groups from the 2015 Carnegie Classification System: Research [147], Doctoral [30], Masters [269], and Baccalaureate [135]. The final institution list contained three Research institutions, four Doctoral institutions, four Masters institutions, and five Baccalaureate institutions. Due to the smaller size of these types of institutions, it was necessary to oversample from the Doctoral, Masters, and Baccalaureate groups.

The final sampling frame consisted of every department chair or school director from the sixteen sampled institutions (approximately 450). Each potential participant was sent a recruitment email outlining the purpose of the study and a link to on online survey where they could register to participate in the study. Overall, forty-one unit heads filled out the recruitment survey (9%). Each of those forty-one participants were mailed a research packet containing two copies of an informed consent form (for IRB purposes), an instruction sheet, and the survey materials from the original FOCS Study. After participants returned one signed copy of the informed consent (in a pre-paid envelope that was included in the packet of materials), they were contacted to schedule a phone interview. The final sample consists of twenty-one chairs and directors across multiple colleges from the sixteen sampled universities (Table 1).

Table 1. Participants.

ID	**CC**	**College**	**Title**	**Sex**	**Time as Unit Head (years)**	**Time in Dept. (years)**	**Interview Length**
1	Research	Education and Human Development	Chair	F	5	14	(01:24:05)
2	Research	Arts and Sciences	Chair	M	5.5	16	(00:50:20)
3	Research	Arts and Sciences	Chair	M	2.5	15	(00:22:25)
4	Bacc.	Business, Humanities, and Social Sciences	Chair	F	3.5	20	(01:02:56)
5	Research	Division of Health Sciences	Director	M	3	3	(00:31:48)
6	Doctoral	Other	Chair	F	7	16	(00:54:00)
7	Bacc.	Liberal Arts	Interim Chair	M	2.5	22.5	(00:33:43)
8	Masters	Humanities and Social Sciences	Acting Chair	M	0.5	8	(01:10:43)
9	Masters	Other	Chair	F	5	18	(00:43:10)
10	Masters	Other	Co-Chair	M	4.5	5	(01:43:53)
11	Bacc.	Health	Chair	M	16	17	(01:06:35)
12	Research	Musical Arts	Chair	M	9.5	15	(01:16:45)
13	Doctoral	Arts and Sciences	Director	F	15	n/a	(00:26:11)
14	Research	Education and Human Development	Chair	F	3.5	22.5	(00:51:59)
15	Research	Arts and Sciences	Chair	M	1.5	18	(01:48:52)
16	Research	Arts and Sciences	Chair	M	0.5	15.5	(01:02:55)
17	Doctoral	Communication, Fine Arts, and Media	Director	M	0.5	0.5	(01:37:32)
18	Research	Arts and Sciences	Chair	M	11	23	(00:53:20)
19	Doctoral	Social Sciences and Communication	Chair	M	11.5	11.5	(01:14:22)
20	Masters	Social and Behavioral Sciences	Chair	F	5.5	24	(01:12:03)
21	Doctoral	Information, Science, and Technology	Chair	M	3	21	(00:55:36)

Overall, the twenty-one unit heads come with a combined 116.5 years of experience in their time as the unit head (average = 5.5 years). While the final sample does represent a variety of colleges, the participants tend to be overrepresented from “Arts and Sciences” colleges. In order to protect the confidentiality of the participants, specific disciplines were not included in Table 1, but a preliminary analysis did not reveal any noticeable differences between disciplines.

The data for this analysis come from semi-structured open-ended interviews with each participant. Because this study was largely exploratory, the research team determined that a qualitative approach would be the best methodological approach. An open-ended interview format was used in order to encourage the participants to thoughtfully engage with the materials presented to them and to allow them to speak freely about potential benefits and consequences of developing a study of faculty activity outside of the classroom. The interviews followed a basic interview guide that contained questions regarding seven topic areas: (1) becoming the unit head; (2) measuring faculty activity in their units; (3) the perceived usefulness of the FACT Study; (4) the state of higher education and criticisms of faculty activity; (5) apprehensions about data collection and the FACT Study; (6) an analysis of the FOCS materials; and (7) supplemental questions about the Delaware Cost Study. The interviews ranged from twenty-two minutes to an hour and forty-three minutes (median = one hour and two minutes) (see Table 1). Between the twenty-one interviews, over twenty-one hours and 30 minutes of interviews were recorded.

The data from the interviews were analyzed using a grounded theory approach, which allows researchers to systematically build theory that is "grounded" in the data (Glaser and Strauss, 1967). This approach, which is frequently used in the social sciences, is appropriate for inductive research. The data were analyzed continually throughout the interview process. Earlier interviews were used to establish conceptual categories that could then be introduced and probed in subsequent interviews. This processed allowed the categories to be modified and refined while drawing connections between them. Eventually, these connections were used to develop the conceptual model necessary to move forward with the FACT Study.

Findings

An in-depth analysis of the interview data revealed three primary conceptual categories. Additionally, within each of the primary categories more detailed and specific findings emerged. The first category of data consists of narratives constructed by participants concerning their roles as faculty administrators. This includes discussions about how they learned to function in their role, expectations about how they measure and evaluate their faculty's activity, and concerns over the amount of administrative duties that they already have. The second category of analysis covers the perceived usefulness of the FACT Study. Specifically, this section is concerned with cataloging real, measurable ways that unit heads could make use of a study that measures faculty activity outside of the classroom. The final conceptual category builds off of the findings from the previous category. While almost all of the participants reported important uses for the FACT Study, those uses often came with a set of apprehensions about the study and data collection in general. While the specific findings from each of the conceptual categories alone offer an insight into the complexities of measuring faculty activity, the findings also have important implications for the theoretical and methodological foundations of the FACT Study.

Administrative Duties of the Unit Head

A discussion on the administrative duties of the unit heads sets an important precedent for how the FACT Study would fit into the set of administrative duties with which unit heads are already tasked. Ideally, data collection for the FACT Study at each participating institution would be coordinated through an institutional research office. However, for a study that is based on faculty activity outside of the classroom, much of the individual units of data will need to be reported by individual faculty members, and ensuring the successful completion of this process would fall on the chairs or directors of each academic unit within the institution. Thus, it is important to understand how such a process would actually work.

Measuring and evaluating faculty activity. Across the board, participants reported that their institutions already have a system in place that measures and evaluates faculty activity; however, the depth and usefulness of these systems varied greatly. Ultimately, because the only formal systems in place tend to be those used in promotion and tenure decisions, certain types of faculty activity end up not getting counted. These activities include the smaller or “invisible” work that often doesn’t make it onto a faculty member’s CV (e.g. informal service, student advising, etc.). While one institution in the sample did use a merit system to recognize those additional activities, this practice was not consistent across all participants, despite the recognition that these types of activities truly make a department run properly.

Another major concern about measuring faculty activity for some participants is that when they are asked to compile and submit reports about their faculty’s activity, they very rarely get extensive or useful feedback from the administration (i.e. Deans or Provosts). Instead, the paperwork tends to “disappear.” One participant explains the precarious nature of administrative work:

> You know, both the administration and the CBA [Collective Bargaining Agreement] think they want all these forms, [they] think they want all this oversight, but then once they design the forms, the forms are completely toothless. They’re afraid of the forms. And so, we just do a lot of meaningless paperwork. (Chair, College of Arts and Sciences, Research Institution [ID_2])

Additionally, very few unit heads had a formal system in place that created and confirmed expectations for their faculty at the beginning of the year. While collective bargaining agreements dictated teaching load and/or workload allotments for individual faculty, rarely did a unit head have a way to discuss the success of his or her department or school as a whole, cohesive unit.

Excessive administrative duties. Many participants were also concerned with measuring their own activities. Specifically, they expressed the feeling of being overwhelmed by the drastic increase over the past few years of the amount of administrative duties that they had to perform. The same unit head from the above quote elaborated on his role as an administrator:

> I'm a better chair now, but, but I'm also a, I've just been a terrible professor the last couple of years. I've become a terrible scholar, and I'll probably eventually go back to being a scholar, but, you know, my career's being destroyed right now, my scholarly career, by a combination of really important stuff, and that's fine with me, but also, really just, you know, stuff that's just a waste of time. (Chair; College of Arts and Science; Research Institution [ID_2])

This quote is indicative of the concern amongst many participants that they are performing an exorbitant amount of administrative duties. This concern also led to the discovery of an additional finding. Unit heads are extremely curious about the extent of their own activity, especially as it compares to the activity of other chairs and directors. Another participant explained:

> [I]t would also be helpful to see what other administrators across the states are doing as well (the department chairs). What is their percentage of time teaching? My percentage time in teaching is 50% and then administration 50%. That's what is assigned. That doesn't always happen. You know that. That doesn't always happen. We end up taking

on more than that a lot of times, and so it would be interesting to see how other chairs are faring as well. (Chair; “Other”; Doctoral Institution [ID_6])

Implications. These discussions reveal various implications about the administrative duties that unit heads face. First, unit heads recognize a lack of efficient ways to measure faculty activity, especially those activities that occur on a smaller scale or that are somewhat “invisible” to more formal methods of evaluation. They are eager for these types of measurements. Thus, any new study of faculty activity needs to prioritize this type of activity. Second, unit heads are inundated with paperwork. If they are going to participate in a new study of faculty activity, two things need to happen: (1) the study needs to only require a low-level time commitment, and (2) if unit heads are going to take the time to fill out more forms, the data and analyses resulting from those forms need to be useful to them. If administrators are going to be using the data to inform the decisions being made about individual academic units, then the unit heads should have the tools necessary to be a part of that process. Additionally, in order to achieve a low-level time commitment, the study needs to utilize methods of data mining to access already-existing data that can prepopulate any metrics included in the study. One final implication from this conceptual category is that unit heads are extremely curious about their own activity and the activity of other chairs and director, both within their own institutions and throughout their discipline. This curiosity is a clear indicator that the FACT Study should include unique supplemental metric that will measure the out-of-classroom activity of unit chairs, which will allow our participants to develop an all-around better understanding of the administrative work of unit heads.

Perceived Usefulness of the FACT Study

The second conceptual category that emerged out of the data is concerned with the perceived usefulness of the FACT Study. Participants were asked a variant of the following question in the interview: “If my office were to develop a national benchmarking study that provided a discipline-level analysis of faculty activity outside of the classroom, would such a study be useful to you? If so, how?” Overall, the participants expressed resounding support for the development of the FACT Study. It is important to clarify that while these data confirm a measurable need for the FACT Study, they are not able to speak to the *extent* of that need. Thus, these results cannot confirm that the FACT Study would be useful for every academic unit head; however, it would offer a specific set of tools that most of the unit heads in this study are eager to use. The reasons for its usefulness are outlined in the following sections.

Developing a deeper understanding. The first way that unit heads would find the FACT Study useful is that it could provide them with a tool to develop a deeper understanding of both their faculty’s activity and the productivity of the unit as a whole. Specifically, they are eager for a way to support and encourage their faculty, rewarding those who perform at high levels and mentoring those who may be underperforming. One participant explained:

> I have always been looking for these types of data because it helps me make a case for whatever, you know, in terms of if I want resources. I am pretty much a data oriented person. I like to show them that, ‘ok, this is how this works,’ but if I have something like this, I can really support my faculty more than just telling them that they are doing a good job. (Director; Division of Health Sciences; Research Institution [ID_5])

An additional layer of support also includes being able to provide data to back up the claims that they make about their faculty, which “would take the discussion of these issues out of just the realm of anecdote and impressions” (Co-Chair; “Other”; Masters Institution [ID_10]). While unit heads tend to have a very in-depth understanding of what their faculty are doing, anecdotes are typically not sufficient evidence of faculty productivity. Instead, they are interested in having a way of “proving” the validity of those claims to their administrators.

Peer comparisons. The participants in this study also remarked on how the FACT Study would allow them to compare themselves, and their unit’s activity, to similar disciplines at peer institutions. This would provide an advanced analytical tool that is currently not available to them. One participant remarked how “it’s always good to see how you stack up against other institutions, and [to see] some things that you might want to begin to take a look at in your program and begin to put in place” (Chair; “Other”; Doctoral Institution [ID_6]). Additionally, a few of the participants expressed interest in having a way to measure themselves up against programs to which they aspire. This is especially useful for institutions or programs that are trying to change their Carnegie Classification. The FACT Study would allow them to gauge their productivity in relation to both peer and aspirant programs.

Getting ahead of the discussion. The unit heads in this study expressed one final, but very relevant, use for the FACT Study. For many, it was a way to get ahead of the discussion that is currently permeating the higher education landscape that is calling for more transparency and accountability when measuring, reporting, and legislating faculty activity. One of the unit heads explains the approach that his institution has taken:

> The feeling at the university was that we really ought to get ahead of this conversation so that when those conversations do come up in the legislature, we’re ready with our own story and our own statistics to head off any measures that come from that… to impose requirements on workload allocation and other kinds of productivity measures. So we

> tried to be a little bit preemptive about it. (Chair; College of Information, Science, and Technology; Doctoral Institution [ID_21])

This chair's university recognized that that the conversation in higher education was moving in a specific direction, and they chose to be proactive in creating their own counter-narrative (specifically, about faculty teaching loads). The FACT Study would provide a similar tool for institutions that are looking to get ahead of the discussion regarding faculty activity outside of the classroom.

Implications. The findings from this conceptual category have invaluable implications about the FACT Study. First, this analysis shows a measurable need for a study that measures faculty activity outside of the classroom. Specifically, unit heads want to be able to use this data in real and meaningful ways that support their faculty and promote productivity within their unit. Second, these findings show that unit heads can be more than passive participants in institutional studies. Instead, they can be active and willing participants in both the data collection phase and the analysis phase. Yes, the data that is collected through the FACT Study can and should be used to inform decision making at the higher-up administrative level, but these results also show that unit heads want to be a part of that decision making process. Chairs and directors, especially in conversation with their deans, are making decisions about their units on a much more frequent basis, so the FACT Study should be a useful analytical tool for *everyone*. Additionally, if unit heads can expect an analytical tool that will actually be useful to them, they are more likely to encourage their faculty to participate fully in the data collection process. Ultimately, when faculty buy in to that process, both the data and the analysis are more useful to everyone.

Apprehensions About the FACT Study

As with any analytical study, any perceived usefulness of it should be approached with a certain level of apprehension. The participants in this study raised a considerable amount of concerns over both the FACT Study and data collection in general. While at first these fears and apprehensions may seem misplaced, they are not to be taken lightly. Each concern that was raised by a participant is another opportunity to reconsider and reimagine the methodology of the FACT Study.

It's not useful. While there was resounding support for the FACT Study, a few participants outright stated that the study would not be useful for them. Moving forward, it is important to consider the reasoning behind their responses because they can illuminate potential flaws in the study's design. One participant claimed that because every institution is unique, it is impossible to compare institutions to one another. Another argued that the concept of data-driven decision-making is a sort of myth in higher education:

> When we do that [selecting peer institutions for evaluation], it sort of plays a very minor role, once people have looked at it, in the ultimate decisions. I can't think of an instance where we actually used something directly as a model for something we put in place. (Chair; College of Arts and Sciences; Research Institution [ID_15])

It's missing nuance. For some of the participants that did find it useful, or thought it could be useful but they weren't sure how, one major concern was that most measures of faculty activity miss the nuance of faculty activity because they rely too heavily on outputs without taking into account any other measure. One unit head explained:

> If it's too metric driven, and the metrics, of course, have some serious flaws in a certain way, they don't account for certain things, and if it's too metric driven, the administrators

> will sort of be beholder to these metrics and very focused on just that, and they'll miss out on things that it doesn't capture, and that sort of thing. So I think there's worry. Definitely, I think that's partly a concern. (Acting Chair; College of Humanities and Social Sciences; Masters Institution [ID_8])

Many unit heads are apprehensive about relying solely on one set of measures when faculty activity is such a complex set of activities. Additionally, participants expressed a serious concern over just measuring "outputs" without accounting for any measure of quality of those outputs. While most agreed that there needed to be qualitative component, they were unsure of exactly of what that would look like.

It's "mission creep." Participants were also concerned that participating in the FACT Study could result, or was indicative of, their administration's proclivity to "mission creep." Specifically, they feared that this study could be used by their administration to make decisions and implement changes within a unit without truly understanding how faculty activity works. One participant also explained that there is a general fear of being measured because they're unsure of the consequences of that measurement:

> I imagine that there's resistance from some because they don't like the idea of a big database that administrators can access at all. It sounds weird probably to some people at [the university], but at [my university] there is an unfortunate part of our culture where people don't want… I mean, they're insecure. They don't want people to see. They don't want it all to be out there. (Acting Chair; College of Humanities and Social Sciences; Masters Institution [ID_8])

It's prone to misinterpretation and it's not credible. Finally, the unit heads in this study also raised concerns about both the credibility of the data and the way that the data will be interpreted. One participant summarized these apprehensions by saying, “You do have to be careful because data, as we all know, can be twisted and used in a ton of different ways (Chair; School of Health; Baccalaureate Institution [ID_11]). Ultimately, the concern expressed here is that the people who are using these data to make decisions have the power to manipulate the data in a way that fits their agenda. This fear speaks to the larger culture in academia concerning the sometimes-confrontational relationship between faculty and administration, and it should not be taken lightly. Additionally, these fears will never be quelled if the data collected aren’t credible. In the very first interview conducted for this study, one chair explained, “I’ve had a general concern here for a number of years. You know we talk a lot about making data-driven decisions, but to do that we’ve got to have good, reliable, trustworthy data that we all agree on” (Chair; College of Education and Human Development; Research Institution [ID_1]).

Implications. While it’s easy to dismiss these concerns as futile attempts at resisting change, the fears and apprehensions communicated through this study have very strong implications. Specifically, because of their candid nature, they should serve as the guiding framework for developing the FACT Study. For the participant that was frustrated with previous attempts at data-informed decision-making, there is an expressed importance in designing an end product that actually meets the needs of its end-users. The criticism that measurements of faculty activity aren’t very nuanced emphasizes the need to expand the methodology beyond just a basic metric that measures counts or outputs. While counts can be very informative, they need to be coupled with some measure of quality. Additionally, the FACT Study needs to allow for an entirely qualitative component where unit heads can provide a narrative of their unit’s activity. The simple addition of this component would provide a sense of agency to the chairs and directors who object to being beholden to a purely quantitative metric.

Finally, in order to address the general fear of being measured by the administration, studies of faculty activity, including the FACT Study, should refrain from measuring faculty activity on the individual level. Instead, data should only be aggregated and compared on a discipline-level of analysis. This suggestion is consistent with previous research that suggests that faculty activity should be measured on the department level because the academic department is a cohesive unit where the different types of faculty activity can actually complement one other (Fairweather, 2002). Instead of every faculty member being productive in every area, the unit can be designed so that productivity is divided amongst its members.

With any study, the data need to be credible and they must be interpreted consistently and ethically. While this finding is not unique to the FACT Study, or even with studies of faculty activity in general, it emphasizes the importance of both establishing a statement of best practice and carefully considering the methodology of the study before it reaches the design phase.

Discussion

During the analysis of these interviews, it quickly became apparent that the heads of academic units are, by and large, very in tune with the current landscape of higher education. They are eager for analytical tools that will allow them to monitor and improve upon the traditional processes embedded in academic administration, and they recognize the benefits of data-informed decision making. Ultimately, their top priority is to ensure the optimal functioning of their unit and to accurately represent the endless amounts of work that their faculty members commit to that success. The findings from this analysis show that the Faculty Activity Trifecta Study would be a necessary and useful tool for understanding discipline-level faculty activity outside of the classroom. Given the current landscape of higher education, the study would also be a timely addition to the various analytical tools available to higher education administrators looking to make data-informed decisions.

However, the findings also revealed a set of apprehensions about the FACT Study that must be taken into account when determining the theoretical and methodological framework of the study. Ultimately, because unit heads are already burdened with an excessive amount of administrative duties, the FACT Study must require a low-level time commitment with a high-level analysis. Moving forward, this study must rely heavily on the faculty activity data that already exists in other databases, and it must be a useful analytical tool for all of its potential end-users.

Further analysis of these data, including a fourth conceptual category that emerged, will be used to develop the specific measures of out-of-classroom faculty activity that will be included in the FACT Study. Specifically, that analysis will focus on which types of activity to include on the metrics and how they should be measured in order to reflect not only the activity outputs but also the effort that was committed to creating those outputs.

The breadth of the findings presented here also has important implications about the methodological design that was selected for this research. While the field of institutional research tends to be dominated by quantitative analyses, this research shows that qualitative data can offer an additional layer of understanding not often explored. It is also an invaluable tool for developing new studies because it allows researchers to ground the theoretical framework of the study in the lived experiences of those that are most affected by it. At the very least, qualitative research, as indicated by this study, can help determine whether or not what is being measuring is even what needs to be measured.

References

AAUP. (1996). The politics of intervention: External regulation of academic activities and workloads in public higher education. *Academe, 82*(1), 46-52.

Fairweather, J.S. (2002). The mythologies of faculty productivity: Implications for institutional policy and decision making. *The Journal of Higher Education*, *73*(1), 26-48.

Glaser, B.G, & Strauss, A.L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. London: Weidenfeld and Nicholson.

Isaacs, H.K., & Middaugh, M.F. (2004). *Out-of-Classroom Faculty Activity Study: A discussion of the results, utilizations, and lessons learned*. Paper presented at the 39th Annual Conference of the Society for College and University Planning, Toronto, ON, Canada.

Jacobs, J.A. (2004). The faculty time divide. *Sociological Forum, 19*(1), 3-27.

Levy, D.C. (2012, March 23). Do college professors work hard enough? *The Washington Post*. Retrieved from http://www.washingtonpost.com.

Metra4.com. (2012, January 4). *Joe Biden visits Central Bucks High School in West Doylestown* [video file]. Retrieved from https://www.youtube.com/watch?v=TBEyGgvjmKk.

Middaugh, M.F. (2001). *Understanding faculty productivity: Standards and benchmarks for colleges and universities*. San Francisco: Jossey-Bass.

Miller, M.A. (1994). Pressures to measure faculty work. *New Directions for Institutional Research, 16*(3), 5-14.

Porter, S.R., & Umbach, P.D. (2001). Analyzing faculty workload data using multilevel modeling. *Research in Higher Education, 42*(2), 171-196.

Terpstra, D.E., & Honoree, A.L. (2009). The effects of different teaching, research, and service emphases on individual and organizational outcomes in higher education institutions. *Journal of Education for Business, 84*(3), 169-176.

Wilson, R. (2001, February 2). It's 10 a.m. Do you know where your professors are? *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com.

Yuker, H.E. (1984). *Faculty workload: Research, theory, and interpretation*. Washington, D.C.: The University: Association for the Study of Higher Education.