Association for Institutional Research

PROFESSIONAL FILES | FALL 2016 VOLUME

Supporting quality data and decisions for higher education.

Letter from the Editor

This volume presents fresh solutions to two perennial challenges for institutional researchers: dealing with declining survey response rates and finding meaning in copious amounts of text data.

Association for Institutional Research

In their article *An Alternative Approach: Using Panels to Survey College Students*, Sarraf, Hurtado, Houlemarde, and Wang describe an experiment to compare results from the standard NSSE administration process with one using much smaller samples and multiple short surveys. Do you believe response rates, scale reliabilities, and factor structures can hold up with panel methods? Read their research to find out if it can work for you.

If you've ever neglected open-ended survey comments simply because you didn't know what to do with all that text, Zilvinskis and Michalski provide a lifeline. Their article, *Mining Text Data: Making Sense of What Students Tell Us*, walks us through how to extract information from text data and identify software to assist in the process. Their examples illustrate text mining with different types of written artifacts and may inspire you to tackle your text data with renewed confidence.

Do you have a solution to a vexing IR challenge? Please share it with AIR Professional File!

Sincerely,

Sharron L. Ronco

IN THIS ISSUE...

EDITORS

ISSN 2155-7535

PROFESSIONAL FILE

ARTICLE 138

Association for Institutional Research

AN ALTERNATIVE APPROACH: USING PANELS TO SURVEY COLLEGE STUDENTS

Shimon A. Sarraf
Sarah Hurtado
Mark Houlemarde
Xiaolin Wang

About the Authors
Shimon Sarraf is the National Survey of Student Engagement's assistant director for Survey Operations and Project Services at Indiana University's Center for Postsecondary Research (CPR). Sarah Hurtado, Mark Houlemarde, and Xiaolin Wang are CPR graduate assistants as well as doctoral students at Indiana University's School of Education.

Acknowledgment
This paper was adapted from a longer paper presented at the annual conference of the Association for Institutional Research in Denver, Colorado, May 2015.

Abstract
Eight short surveys based on select items from the National Survey of Student Engagement (NSSE) were administered to approximately five hundred students over a nine-week period at five diverse colleges and universities. The goal of the experiment was to investigate what impact a survey panel data collection approach would have on recruitment, survey data quality indicators, and scale properties. Results indicated higher response rates, shorter survey duration, and minimal impact on scale factor structures. However, both cost of incentives and panel member attrition make this alternative survey method less attractive than it would be otherwise.

INTRODUCTION

Survey researchers in higher education are engaged in an uphill battle with declining student response rates. Student cynicism, survey fatigue, and disinterest pose a substantial threat to optimal survey participation (Porter, 2005). Despite significant challenges with the collection of student feedback, demands for evidence-based decisions in higher education continue to increase (Zhang, 2010). In light of these conflicting trends, identifying potential alternatives to standard survey practices may prove useful for a variety of higher education constituents, including institutional researchers and higher education scholars. Some scholars have already seriously considered this issue. For instance, Stern, Bilgen, and Dillman (2014) have proposed survey panels—a group of individuals that have agreed to respond to multiple future survey completion requests—as a potential solution to declining response rates. Despite today's varied approach to implementing a survey panel, researchers have typically asked a group of individuals the same set of questions at different points in time. Though some academic disciplines, government agencies, and businesses have used survey panels for decades, higher education researchers and administrators have not often employed them to help with their data collection needs (Zhang, 2010).

As a provider of assessment data to hundreds of colleges and universities across North America, and having witnessed its own response rates decline over the past decade, the National Survey of Student Engagement (NSSE) experimented with survey panels at five diverse institutions in spring 2014. The panel administration included eight surveys, with eight to ten items each, administered over nine weeks. Using standard NSSE administration results as a benchmark for each institution, this study documents the impact that using survey panels has on recruitment, survey data quality indicators, and scale reliability and factor structure.

SURVEY PANEL HISTORY

For decades, survey panels have been used in market and medical research (Callegaro et al., 2014; Callegaro & Diogra, 2008). These survey administrations have historically been conducted in person, by phone, and by mail, and have recently transitioned to the Internet (Callegaro et al., 2014). Callegaro et al. estimate that the first online survey panel was administered in the mid-1980s in Europe. About a decade later, this method became popular in the United States. In particular, annual conferences dedicated to public opinion and survey research professionals, such as the one held for the American Association for Public Opinion Research (AAPOR), are rife with presentations from for-profit, not-for-profit, and government organizations (AAPOR, 2015). According to Callegaro et al., the benefits of online survey panels are threefold: (1) quick data collection, (2) low administration cost, and (3) sampling efficiency. With a heavy reliance on online student surveys, these benefits may resonate with the needs of institutional researchers and higher education scholars.

National education datasets developed by the federal government have used survey panels for better understanding the backgrounds and academic experiences of college students and their career trajectories. These include such studies as the Baccalaureate and Beyond and the Beginning Postsecondary Students Longitudinal Study (National Center for Education Statistics [NCES], n.d.). These survey panels are distinct from many of the online survey panels previously discussed because they contact subjects only every few years, and not on a weekly or monthly basis. In terms of college student assessment, few researchers appear to be developing their own survey panels, although some have used them to answer methodological questions related to survey response behavior (Porter & Whitcomb, 2005; Sharkness & Miller, 2014).

Panel Definitions and Design Considerations

Various researchers define survey panels differently, so it is important to clarify definitions. The traditional "panel" definition refers to a longitudinal survey panel that involves asking the same individuals the same questions across different points in time (Callegaro et al., 2014; Goritz, Reinhold, & Batinic, 2000; Hsiao, 2014), with each point in time referred to as a "wave." This approach is inherently suited for studying particular changes among subjects. Many survey panels have evolved into access panels; an access panel is essentially a "database of potential respondents who declare that they will cooperate for future data collection if selected" (International Organization for Standardization [ISO], 2012, p. 1). One form of panel survey utilizes a split questionnaire approach by dividing longer questionnaires into smaller ones (Raghunathan & Grizzle, 1995). These various panel methods have been used in an effort to minimize missing data (Vriens, Wedel, & Sandor, 2001).

Types

There are two primary survey panel types based on how members join: nonprobability panels and probability panels (Callegaro et al., 2014). Nonprobability panels are open and allow anyone to volunteer to become a panel member. When members self-select into a panel, it is not possible to know the probability of selection, hence the nonprobability label. In contrast, individuals cannot join probability panels unless they have been invited to do so. Probability panels require that "all members of the population of interest have a known, non-zero probability of receiving an invitation to join" (Callegaro et al., 2014, p. 7).

Sampling

Segers and Franses (2014) found that it was difficult to survey every panel member at each point in time because of cost and the potential for nonresponse. To address these concerns, they utilized a rotating sampling method whereby panel members participated for a fixed amount of time with new panel members integrated at each wave. Other sampling methods include continuous, time, randomized, and matrix sampling. Continuous sampling means that the researcher surveys all individuals at each wave for the extent of the project. Time sampling refers to panel members being surveyed on a regular basis (e.g., biweekly) but not for every wave; they are rotated into waves, helping to ensure data are collected weekly, for instance, or on some other predetermined schedule. Randomized sampling refers to panel members being chosen at random for each particular wave. Finally, matrix sampling calls for a panel survey to be divided into groups of questions with panel members answering only one question group (Segers & Franses, 2014).

Incentives

Incentives are frequently used in survey research in an effort to increase response rates, and panel studies are no exception. There are various types of incentives (e.g., guaranteed prepaid, guaranteed postpaid, and lottery). In relation to panel studies, there is conflicting evidence regarding the different incentive types and their effectiveness. Goritz (2006) found that cash lotteries (both one large prize and multiple smaller prizes) did not reliably increase panel response or retention rates. In a study on the impact of individual payment on a three-wave longitudinal experiment, Goritz, Wolff, and Goldstein (2008) found that guaranteed payments had a negative effect on the first wave and a positive effect on the second wave. Additionally, after a review of the literature, Callegaro et al. (2014) found that postpaid incentives are most impactful when the members have prior experience or knowledge of the organization administering the survey. None of these studies used a college student population so the current study should shed light on incentive effectiveness, and, more specifically, on guaranteed prepaid incentives.

This study was based on a probability panel that used a split-questionnaire design with continuous sampling, and both guaranteed prepaid and lottery incentives.

STUDY RATIONALE AND RESEARCH QUESTIONS

Though aggregate response rates for NSSE have slowly declined for years (the 2014 administration being a recent exception), they are still high enough to reliably estimate institution-level engagement for the vast majority of participating schools. As a recent study using NSSE data has shown, low response rate estimates of student engagement are very similar to estimates based on high response rates (Fosnacht, Sarraf, Howe, & Peck, in press). Nevertheless, it is incumbent on NSSE and other higher education researchers to look closely at alternative data collection methods to see if there are more-effective ways to collect student feedback. Whether or not minimal response rates and respondent counts can estimate institution-level engagement sufficiently, low response rates matter to skeptical audiences who believe higher response rates are required for decision making. If response rates drop precipitously for a significant number of institutions in the future, those interested in student opinion may need to rely more heavily on other data collection methods, such as focus groups, nonprobability sample surveys, and/or survey panels to supplement their standard survey administrations.

The following research questions guide this study and help us assess whether survey panels can serve NSSE and the wider institutional research community. Questions are grouped into three general areas for assessing the viability of panels: (1) recruitment, (2) survey data quality indicators, and (3) scale reliability and factor structure.

Recruitment

1. What percentage of students tried to register for the NSSE survey panel, and how do these rates compare to standard administration invitation response rates?
2. What were claim rates for the panel's guaranteed incentives?
3. How do panel members compare demographically to non-panel members and to standard administration respondents?

Survey Data Quality Indicators

4. What are response rates to individual panel surveys, and how do they compare to standard administration rates?
5. What are survey duration and completion results by administration type?

Scale Reliability and Factor Structure

6. Do panel and standard survey administration methods produce similar scale scores?
7. Does the factor structure of scales vary by survey administration method?

Study Background

Data Source and Sample

To address the research questions, we combined NSSE standard administration survey data with panel data for five colleges and universities. These institutions differed by size (total undergraduate enrollment), status (public or private), and designation (college or university).[1] To preserve the anonymity of the five study institutions, we have named them Small Private College, Small Private University, Medium Public University, Medium Private University, and Large Public University.

Each institution participated in the 2013 standard administration with the exception of the Large Public University; that university's data came from an experimental 2014 standard administration that used a smartphone-optimized version of NSSE. All five institutions participated in a spring 2014 administration using survey panels. In fall 2013 NSSE staff contacted 12 diverse colleges and universities to offer them an experimental, no-cost survey panel administration during the following semester (spring 2014). NSSE staff selected five of the six institutions that expressed interest.

Because of other survey commitments at the Large Public University, NSSE staff randomly selected 25% of all first-year and senior students to participate in either the panel or smartphone experiments, and then followed up by randomly assigning the selected students to either the panel or the standard administrations. The other four institutions provided all first-year student and senior records for sampling. Though not part of the original research design, the concurrent panel and standard administrations for the Large Public University allow for stronger claims about panels, whereas at the other institutions we compare results obtained at different times and from different student populations.

We analyzed 3,331 respondents from both administration types: 2,832 standard administration and 499 panel respondents. Of that total, 67% were female, 17% were underrepresented minority, and 6% were part-time students. With the exception of sex, these demographics reflect missing data for some institutions. We also used 12,272 first-year student and senior non-panel members from the 2014 panel administration to assess recruitment success and panel composition; this included 1,900 students at the Large Public University assigned to the experiment, which represents a fraction of all its first-year students and seniors.

Panel Administration Details

Our goal was to recruit 50 first-year students and 50 seniors from each institution. We drew repeated random student samples by institution and class level in order to send panel registration invitation e-mails; we sent registration invitations to 6,595 students, ranging from 650 to 1,950 per institution. Invitations emphasized that each of the eight surveys would take about one minute to complete. We sent students only one panel registration e-mail and concluded the registration process over about two days. Those that attempted to register after all one hundred openings had been filled were told that it would no longer be possible to join the project. As promised in the registration invitation e-mail, all panel members could immediately retrieve a $10 Amazon.com gift card using an online portal that was created for the study. Additionally, the registration e-mail informed students that the names of students who completed all eight panel surveys would be included in a drawing for one $250 Amazon.com gift card at each school.

Upon registration, students received an e-mail requesting they complete the first survey, followed by a reminder e-mail two or three days later if they had not responded. Approximately every week we delivered another e-mail invitation and follow-up reminder for another panel survey. As the administration for each survey began, we also posted the survey link to the online portal. It took approximately nine weeks to administer all eight surveys. During the last week, a final reminder was delivered to students who had not completed all eight surveys, encouraging them to log in to the online portal to complete all surveys.

Standard Administration Details

As part of a standard NSSE administration, all first-year and senior students received survey recruitment messages sometime between February and May, and the survey officially closed on June 1. As with all institutions, NSSE sent students at the five study schools five e-mail recruitment messages. Institutions decided to use survey incentives to boost response rates and, when appropriate, they included related text in all recruitment messages. Each recruitment e-mail included a URL that linked to the online survey and to

[1] *The labels "small," "medium," and "large" signify that total undergraduate enrollment is either fewer than 5,000 (small), between 5,000 and 15,000 (medium), or more than 15,000 (large).*

Table 1. NSSE Panel Survey Content

Survey #	Topics	Survey Item Count
1	Collaborative Learning, Student–Faculty Interaction	8
2	Reflective & Integrative Learning, Academic Major, Class Level	9
3	Higher-Order Learning, Writing Practices, Age	8
4	Quantitative Reasoning, Effective Teaching Practices, Parental Education	9
5	Discussions with Diverse Others, Learning Strategies, Course Challenge, Living Location	9
6	High-Impact Practices, Class Preparation, Reading	9
7	Quality of Interactions, Number of Courses Taken (overall and online), Race	8
8	Supportive Environment, Other Institution Experience	10

the informed consent statement. The informed consent statement advised students that the survey would take a total of between 15 and 18 minutes to complete. Though we did not use available data for this study, institutions could append topical module item sets ranging in content from academic advising to civic engagement, and participate in a consortium that administered additional items.

Survey Content

The 2013 standard administration survey included approximately 104 survey items.[2] The eight experimental panel surveys, however, had 70 items combined (see table 1). The panel administration incorporated arguably the most important topics from the standard instrument, including survey items from 10 primary scales used for official NSSE reporting and several other important student experiences, background, and demographic items, including academic major and parental education. Panel survey content order did not follow the standard core instrument's layout. For a complete list of panel survey items, readers may contact the authors at nsse@indiana.edu.

STUDY FINDINGS

Recruitment

1. What percentage of students tried to register for the NSSE survey panel, and how do these rates compare to standard administration invitation response rates?

After a single e-mail invitation, between 6% and 20% of invited students registered, or attempted to register, for the panel administration at the five participating institutions (see table 2). The proportion that responded varied significantly by institution, overall, and by class level ($p < .001$). At all but one institution, a lower proportion of first-year students than seniors registered for the panel study. Differences in registration rates favored seniors by 2 to 8 percentage points. Compared to the standard administration response to the first recruitment message, interest in the panel was appreciably higher at all five institutions. The Small Private College showed the greatest difference with 8 percentage points (12% standard administration rate versus 20% panel rate), while the Small Private University and the Medium Private University showed the least, with a 2 percentage point difference at each.

---

[2] *For a survey facsimile, see NSSE (n.d.b).*

Table 2. Initial Response to Participation Request by Administration Type

		Panel Registration[b]				
	Standard Administration[a]	Overall	First-Year Students (FY)	Seniors (SR)	FY vs. SR (X^2)	Panel Gift Card Retrieval Rate
Small Private College	12.0%	20.0%	17.4%	23.0%	+	78.0%
Small Private Univ.	10.0%	12.4%	9.4%	17.4%	***	75.0%
Medium Public Univ.	2.0%	6.9%	6.0%	8.0%	+	78.0%
Medium Private Univ.[c]	4%	6.0%	6.9%	5.4%		79.8%
Large Public Univ.	2.0%	9.3%	7.0%	11.4%	**	81.0%
X^2	n/a	***	***	***		

Note: + p < .1; * p < .05; ** p < .01; *** p < .001 (X^2-*test*)

[a] Standard administration results reflect response rate five days after initial invitation to complete survey.

[b] Panel registration results reflect total registered panel members and unsuccessful registration attempts divided by the total number of registration invitations delivered. Approximately 105 students across all five schools tried to register but were not included in the panel because of the 100 student quota per institution.

[c] This institution offered an incentive for completing the standard administration survey.

2. What were claim rates for the panel's guaranteed incentives?
Overall, 78% of the 499 panelists claimed their $10 Amazon.com gift card. There was no statistically significant difference in claim rates among institutions (table 2). Claim rates were also unrelated to class level: 76.2% of first-years and 80.4% of seniors claimed their incentives.

3. How do panel members compare demographically to non-panel members and standard survey administration respondents?
Panel members appeared very similar to non-panel members at each institution for two of the three demographic variables we analyzed: full-time enrollment and underrepresented minority status (see table 3). We found no statistically significant differences for these two demographic variables at the five institutions. However, for four of the five (the Large Public University being the exception), there was a 10 to 17 percentage point greater proportion of females among panel members compared to non-panel members.

In terms of these demographic variables, panel members were generally comparable to standard administration respondents at each institution, as well. We see a single statistically significant difference in the proportion of females by administration type across the five schools: 51% of the Large Public University's panel members were female whereas 62% of its standard administration respondents were female.

Survey Data Quality Indicators

4. What are response rates to individual panel surveys, and how do they compare to standard administration rates?
Panel survey response rates at the five institutions ranged from 95% for the Small Private College's first survey to 71% for the Medium Private University's seventh survey (table 4). All institutions

Table 3. Panel Member Characteristics Compared to Non-Panel Members and Standard Administration Respondents by Institution

		Female	Sig.	Full-Time	Sig.	Minority	Sig.
Small Private College	Panel member	74%		99%		11%	
	Non-panel member	57%	*	99%		10%	
	Standard respondent	78%		100%		9%	
Small Private Univ.	Panel member	70%		97%		n/a	
	Non-panel member	57%	*	97%		n/a	
	Standard respondent	62%		97%		5%	
Medium Public Univ.	Panel member	63%		94%		18%	
	Non-panel member	53%	*	90%		19%	
	Standard respondent	65%		91%		18%	
Medium Private Univ.	Panel member	75%		91%		25%	
	Non-panel member	60%	*	89%		28%	
	Standard respondent	71%		93%		24%	
Large Public Univ.	Panel member	51%		n/a		n/a	
	Non-panel member	49%		n/a		n/a	
	Standard respondent	62%	*	n/a		n/a	

Note: * $p < .05$ (X^2-*test*)

had response rates over 90% for the first survey; rates for the eighth survey ranged from 72% to 86%. With the exception of the second survey, statistical tests indicate that response rates to individual panel surveys do not vary across institutions. As is common with panel administrations, response rates declined across the eight panel surveys. Percentage drops in response rates from the first to the eighth survey ranged from 11% for the Small Private College to 21% for the Medium Private University. Standard administration response rates ranged from 34% for the Small Private College to 12% for the Large Public University. In all cases, response rates for panel surveys far exceeded final rates for corresponding standard administrations at each institution.

5. What are survey duration and completion results by administration type?

Analysis showed that all panel surveys averaged a total of 9.2 minutes to complete, compared to 11.7 minutes for standard survey core items. If a panel member completed one item in a survey, he or she almost always completed all the items.

Table 4. Response Rates by Administration Type

	Panel Surveys									Standard Administration	
	1	2	3	4	5	6	7	8	Panel Members (n)	Response Rate	Respondents (n)
Small Private College	95%	93%	85%	85%	84%	84%	84%	86%	100	34%	333
Small Private Univ.	92%	83%	80%	77%	78%	78%	78%	78%	100	28%	341
Medium Public Univ.	92%	88%	84%	83%	81%	80%	79%	76%	100	15%	753
Medium Private Univ.	93%	78%	79%	74%	72%	73%	71%	72%	99	21%[a]	718
Large Public Univ.	91%	83%	81%	80%	77%	78%	77%	75%	100	12%[b]	308
X^2		*								n/a	

Note: * p < .05 (X^2-*test*)

[a] Institution offered an incentive for responding to the standard administration survey invitation.

[b] Institution offered incentive after initial invitation; a random sample of students was used for this school's administration.

Table 5. Average Number of Panel Surveys Completed and Survey Completion Rates by Administration Type

		Completion Rate[a]		
	Panel Surveys Completed (of 8)	Panel	Standard Administration	Sig.
Small Private College	7.1	81%	73%	
Small Private Univ.	6.8	73%	81%	+
Medium Public Univ.	6.7	76%	77%	
Medium Private Univ.	6.3	66%	80%	**
Large Public Univ.	5.6	72%	n/a	
X^2	**	–	–	

Note: + p < .1; ** p < .01 (X^2-*test*)

[a] Completion rate defined as the percentage of respondents completing 90% of all survey items.

However, not all panel members completed every survey. Panel members completed an average of 6.5 out of 8 surveys with notable variation across institutions (table 5). Students at the Large Public University completed 5.6 surveys on average, whereas those at the Small Private College completed 7.1.

Completion rates, defined as the proportion of students completing at least 90% of standard NSSE items or all survey panel items combined, favored the standard administrations of the Small Private University and the Medium Private University by 8 and 14 percentage points, respectively. In contrast, the Small Private College and the Medium Public University showed no statistically significant differences. We could not reliably calculate the Large Public University completion rates for the mobile-optimized experimental administration due to a programming error.

Scale Reliability and Factor Structure

Knowing that a panel administration approach would not substantially change psychometric scale properties is very important to assessing overall panel viability. In order to shed light on this issue, we answered two questions using NSSE scales, otherwise known as engagement indicators (EIs).[3] EI scores measure the frequency with which students engage in various educationally enriching behaviors, students' perceptions of campus support, and the quality of students' interactions with different groups, such as faculty and other students.

6. Do panel and standard survey administration methods produce similar scale scores?

To assess the reliability of EI scores, we compared these scores by administration type for both first-year students and seniors. Differences in EI scores might be attributed to the fewer number of panel respondents relative to the standard administration, or to unknown changes that occurred at campuses in the one-year lapse between administrations, although we would not expect meaningful differences unless an institution undertook major programmatic changes. To determine whether EI scores differed by administration type, we first calculated effect sizes by subtracting the standard administration EI score from the panel EI score and dividing the difference by the standard administration EI standard deviation. We then reported and evaluated the absolute values of the effect sizes by using guidelines developed by NSSE staff suggesting that an effect size of .1 is considered small, .3 is considered medium, and .5 is considered large (Rocconi & Gonyea, 2015). We used t-tests to determine whether any differences were statistically significant at the .1 alpha level after applying a correction for false discovery rates, which is a concern when conducting multiple tests concurrently (Benjamini & Hochberg, 1995).

Our study suggests that both survey administration approaches generally yield similar results (table 6). Out of the 100 comparisons, only nine showed both meaningful and statistically significant differences, ranging from .36 to .81 effect size. Most notably, the Large Public University, which had the methodological advantage of concurrent random assignment into either the panel or standard administration, showed no differences. For the 20 comparisons for each institution, we see anywhere between zero and three notable differences. For senior populations at the Small Private University and the Medium Public University we found two notable differences, which represents the greatest class-specific discrepancy for any institution. In terms of certain scales that may present particular challenges, Quantitative Reasoning appears to have the greatest difference between the two survey administration approaches.

7. Does the factor structure of scales vary by survey administration method?

We used multi-group confirmatory factor analysis to test for measurement invariance (or consistent factor structure) by the two types of survey administration methods. Confirming measurement invariance ensures that scores relate "to the same set of observations in the same way in each group" (Borsboom, 2006), which allows the researcher to reliably draw conclusions from intergroup comparisons. Based on multiple

[3] The 10 EIs analyzed are (1) Higher-Order Learning, (2) Reflective and Integrative Learning, (3) Learning Strategies, (4) Quantitative Reasoning, (5) Collaborative Learning, (6) Discussions with Diverse Others, (7) Student–Faculty Interaction, (8) Effective Teaching Practices, (9) Quality of Interactions, and (10) Supportive Environment. For more information about EIs, see NSSE (n.d.a).

Table 6. NSSE Engagement Indicator Effect Sizes Comparing Panel and Standard Administration Scores

Class Level	Engagement Indicator	Small Private College	Small Private Univ.	Medium Public Univ.	Medium Private Univ.	Large Public Univ.
First-Year	Higher-Order Learning	0.05	0.55 *	0.01	0.28	0.01
	Reflective & Integrative Learning	0.15	0.28	0.18	0.16	0.08
	Learning Strategies	0.11	0.29	0.11	0.29	0.28
	Quantitative Reasoning	0.81 *	0.48	0.29	0.01	0.31
	Collaborative Learning	0.10	0.11	0.09	0.13	0.14
	Discussions with Diverse Others	0.24	0.07	0.13	0.23	0.22
	Student–Faculty Interaction	0.14	0.07	0.34	0.14	0.44
	Effective Teaching Practices	0.20	0.02	0.22	0.22	0.16
	Quality of Interactions	0.00	0.00	0.29	0.03	0.14
	Supportive Environment	0.04	0.01	0.45 *	0.14	0.05
Senior	Higher-Order Learning	0.01	0.03	0.28	0.24	0.01
	Reflective & Integrative Learning	0.08	0.02	0.19	0.06	0.19
	Learning Strategies	0.28	0.15	0.25	0.30	0.14
	Quantitative Reasoning	0.31	0.33	0.36 *	0.42 *	0.02
	Collaborative Learning	0.14	0.46 *	0.00	0.14	0.16
	Discussions with Diverse Others	0.22	0.03	0.08	0.21	0.01
	Student–Faculty Interaction	0.44 *	0.10	0.18	0.00	0.09
	Effective Teaching Practices	0.16	0.24	0.32	0.30	0.21
	Quality of Interactions	0.14	0.53 *	0.09	0.10	0.09
	Supportive Environment	0.05	0.09	0.52 *	0.22	0.13

Note: * p < .1 (*t*-test)

confirmatory factor analysis model results, we grouped each of the 10 scales for both class levels into one of five increasingly invariant (or consistent factor structure) categories: (1) variant, (2) configural invariance, (3) weak factorial invariance, (4) strong factorial invariance, and (5) strict factorial invariance.[4] In order for NSSE to use survey panels for measuring any specific EI score in the future (and to compare these results to standard survey administration results), strong or strict factorial invariance for each scale by class level would be needed.[5]

Our results indicated that 8 out of 10 first-year student EIs met the criteria for strict factorial invariance. Higher-Order Learning and Supportive Environment met the lesser criteria for strong factorial invariance. Senior year results indicated that five EIs had strict factorial invariance while three had strong factorial invariance. Higher-Order Learning and Supportive Environment results indicated variant factor structures between panel and standard administrations, which would make any inter-administration comparisons untenable.

LIMITATIONS, DISCUSSION, AND CONCLUSIONS

Limitations

There are several limitations associated with this study's design and analyses. First, results for four of the five study institutions came from comparing potentially different student populations during the spring 2013 and spring 2014 semesters, which could have unknown effects on study outcomes. Second, we guaranteed that all panel members would receive a $10 Amazon.com gift card incentive as well as a chance to win a $250 Amazon.com gift card, whereas standard administration sample members at three institutions did not get any incentive and the other two conducted lotteries. Making any firm conclusions about the relative attractiveness of panels is therefore confounded; the incentives could have impacted data quality outcomes, especially because they relate to survey completion. Related to this point is whether any selection bias exists given that panel members chose to join the panel after just one registration e-mail. Although we reviewed several key demographics and determined that panel members were generally representative, they may still be different in unknown ways that influence engagement results. Finally, given the relatively small number of panel respondents at each institution, we could not analyze results by subgroups. This study's results may not offer sufficient information for those who typically review survey results by academic major fields or other subpopulations.

Discussion

This study provides various insights about college student survey panels, one possible alternative to the standard survey administration approach used by NSSE, other national survey projects, and many institutional research offices. Foremost among our findings, panel-registered students from all five study institutions responded to panel surveys at very high rates (70% or more), far above response rates for corresponding standard administrations. These findings support the idea of Stern et al. (2014) suggesting survey panels may be one solution for declining response rates.

As social exchange theory might suggest, these high response rates likely stem from the minimal cost of participation (cost for each survey was about a minute answering relatively innocuous questions) in relation to several perceived benefits, including the incentives offered and the opportunity for students to provide helpful feedback about their college experience. Another valid explanation is that students may feel a strong ethical obligation to participate

---

[4] To accomplish this categorization, models were run separately for each administration type until the same model fit both groups well. If no model fit both groups, we rejected measurement invariance and pursued no additional testing. Assuming a model fit both groups well, we then ran tests for the four types of invariance sequentially. Once a lower level of invariance was tested and rejected, we did not proceed with running tests for higher levels of invariance. Criteria used for determining acceptable model fit were RMSEA < .06, Chi-square p-value > .05, and CFI/TLI > .90. Strict factorial invariance required chi-square difference test p-values greater than .05 and ΔCFI values of less than .01.

[5] For a detailed review of multi-group confirmatory factor analysis methods, see Little and Slegers (2005).

after receiving their $10 gift card. A guaranteed incentive of this amount is generally unheard of among NSSE standard administrations and may not be a particularly realistic amount for some institutions to offer students, especially if they want more than 100 panel members. For this reason, we encourage others to experiment with lowering the incentive's dollar value to minimize expenditures but without sacrificing too much on response rates. The survey methodology literature on college student panel incentives is thin, if not entirely missing, so further investigation is certainly warranted.

With the exception of gender, we also found panel members to be similar to non-panel members and standard administration respondents using full-time enrollment and underrepresented minority status. Females traditionally respond to NSSE survey requests at higher rates, so their overrepresentation among panel members is not especially surprising. These results bode well for the use of panel studies since administrators may not have to worry excessively about self-selection bias.

Consistent with similar studies completed in market and medical fields, results indicate, unsurprisingly, that panel studies do offer a rapid way to collect information from students. Panel members answered each of the eight- to ten-item surveys in about a minute with nearly zero item skipping or nonresponse. The short duration and practically nonexistent item nonresponse are other reasons that the panel survey approach may be a viable alternative to standard administration practices. This study's conclusions differ from the conclusion of Apodaca, Lea, and Edwards (1998) that individuals are reluctant to participate in surveys with multiple components.

In terms of NSSE scales, the vast majority of results indicated scores originating from panels are comparable to those from standard administrations, even with a limited number of respondents. With the exception of two EIs for seniors, scale factor structures do not appear to be affected by the data collection method, something that would be very important for analyzing panel and standard administration results longitudinally. The reasons for Higher-Order Learning and Supportive Environment scales failing invariance testing are unclear at this time. At a minimum, this finding would complicate combining results from these two different survey administrations.

In contrast to these favorable panel results, this study also highlights several issues of concern with using survey panels. Based on our recruitment findings, it does not appear that students perceive panel membership to be significantly more attractive than a standard administration approach, even with the $10 guaranteed Amazon.com gift card and promises of a minute-long survey each week. There are several possible explanations, such as students weighing the protracted involvement over nine weeks against the incentives provided and determining that the costs actually outweigh the benefits. Second, we saw that panel member attrition drove overall missing data levels for some institutions to a point that resembles or surpasses standard administration results. Though it is challenging to generalize based on our results, when we look at the percentage of students that complete 90% of all survey items by administration type we can reasonably conclude that completion rates are not always improved by dividing up a longer survey. Variations of this study's panel design, such as rotating panel members to reduce fatigue, may yield more-promising results in terms of missing data levels.

Amidst a growing need for survey data to inform decisions in higher education, students are at increased risk for survey fatigue. Survey panels may offer some relief by limiting the need to administer surveys to entire campus populations. While 100 participants joined the panel and received eight survey invitations, the vast majority of students at each institution did not receive more than a single registration e-mail, thus reducing any potential frustration with multiple requests. High panel response rates may effectively limit the need to send several reminders to sample members. Additionally, we provided students the option to go back and complete previous waves of the survey using an online portal, thereby eliminating the need for several survey reminder messages that could increase perceptions of burden and actual survey fatigue.

Conclusions

College and university students often perceive requests to complete a relatively long 15- to 18-minute survey as burdensome, especially considering the many other survey requests they receive. This overload of surveys challenges survey administrators to think about more-effective ways to collect student feedback than the customary cross-sectional survey design that NSSE and other survey research projects currently employ. Survey panels represent one data collection alternative that is worthy of further investigation: our results point to high response rates, short completion times, and minimal impact on measurement scales. Panel member attrition and associated missing data levels, however, makes this option potentially problematic. The price for encouraging panel membership by using incentives may be cost prohibitive to some as well, especially if the researcher needs significantly more panel members for analyzing campus subpopulations. Our hope is that others will start experimenting with panels to see if there are ways to address these issues. Obviously, we do not know what the future holds for the standard survey administration approach, but investigating some viable alternatives may prove helpful in the future.

REFERENCES

American Association for Public Opinion Research (AAPOR). (2015, May). Conference program from the annual conference of the American Association for Public Opinion Research (AAPOR) held in Hollywood, FL. Retrieved from http://www.aapor.org/AAPOR_Main/media/AM15/AAPOR-15-FP_FNL.pdf

Apodaca, R., Lea, S., & Edwards, B. (1998). *The effect of longitudinal burden on survey participation*. Paper presented at the annual conference of the American Association for Public Opinion Research (AAPOR), St. Louis, MO.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society 57*(1), 289–300.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*, 176–181.

Callegaro, M., Baker, R., Bethlehem, J., Goritz, A. S., Krosnick, J. A., Lavrakas, P. J. (2014). Online panel research: History, concepts, applications a look at the future. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 1–22). Chichester, UK: Wiley & Sons.

Callegaro, M., & Disogra, C. (2008). Computer response metrics for online panes. *Public Opinion Quarterly, 72*(5), 1008–1032.

Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. (in press). How important are high response rates for college surveys? *The Review of Higher Education.*

Goritz, A. S. (2006). Cash lotteries as incentives in online panels. *Social Science Computer Review*, *24*(4), 445–459.

Goritz, A. S., Reinhold, N., & Batinic, B. (2000). Online panels. In B. Batinic, U. Reips, M. Bosnjak (Eds.), *Online social sciences* (pp. 29–51). Gottigen, Germany: Hogrefe & Huber.

Goritz, A. S., Wolff, H., & Goldstein, D. G. (2008). Individual payments as a longer-term incentive in online panels. *Behavioral Research Methods*, *40*(4), 1144–1149.

Hsiao, C. (2014). *Analysis of panel data* (Vol. 54). Cambridge: Cambridge University Press.

International Organization for Standardization (ISO). (2012). ISO 20252 *Market, opinion and social research: Vocabulary and service requirements* (2nd ed.). Geneva: Author.

Little, T. D., & Slegers, D. W. (2005). Factor analysis: Multiple groups. In B. Everitt & D. Howell (Eds.); and D. Rindskopf (Section ed.), *Encyclopedia of statistics in behavioral science* (pp. 617–623). Chichester, UK: Wiley.

National Center for Education Statistics (NCES). n.d. Surveys & programs. Institute of Education Sciences, National Center for Education Statistics, Washington, DC. Retrieved from http://nces.ed.gov/surveys/SurveyGroups.asp?group=2

National Survey of Student Engagement (NSSE). n.d.a. Engagement indicators. Author, Bloomington, IN. Retrieved from nsse.iub.edu/html/engagement_indicators.cfm

[3M]. n.d.b. Survey instrument. Author, Bloomington, IN. Retrieved from nsse.iub.edu/html/survey_instruments.cfm

Porter, S. R. (2005). Survey research policies: An emerging issue for higher education. *New Directions for Institutional Research, 127,* 5–15.

Porter, S. R., & Whitcomb, M. E. (2005). Nonresponse in student surveys: The role of demographics, engagement and personality. *Research in Higher Education*, *46,* 127–152.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90*(429), 54–63.

Rocconi, L., & Gonyea, R. M. (2015). *Contextualizing student engagement effect sizes: An empirical analysis.* Paper presented at the Association for Institutional Research annual conference, Denver, CO.

Segers, R., & Franses, P. H. (2014). Panel design effects on response rates and response quality. *Statistica Neerlandica*, *68*(1), 1–24.

Sharkness, J., & Miller, K. (2014). *Who fills out multiple surveys? Tracking response using online panels.* Paper presented at the Association for Institutional Research annual conference, Orlando, FL.

Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The state of survey methodology challenges, dilemmas, and new frontiers in the era of the tailored design. *Field Methods, 26* (3), 284–301.

Vriens, M., M. Wedel, & Z. Sandor (2001). Split-questionnaire designs: A new tool for survey design and panel management. *Marketing Research, Summer,* 15–19.

Zhang, L. (2010). The use of panel data models in higher education policy studies. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 307–349). New York: Springer.

This page left intentionally blank

PROFESSIONAL FILE

ARTICLE 139

Association for Institutional Research

MINING TEXT DATA: MAKING SENSE OF WHAT STUDENTS TELL US

John Zilvinskis
Greg V. Michalski

About the Authors
John Zilvinskis is a doctoral student in the Higher Education and Student Affairs (HESA) program at Indiana University. Greg V. Michalski is Director of Institutional Analytics and Research, Florida State College at Jacksonville.

Abstract
Text mining presents an efficient means to access the comprehensive amount of data found in written records by converting words into numbers and using algorithms to detect relevant patterns. This article presents the fundamentals of text mining, including an overview of key concepts, prevalent methodologies in this work, and popular software packages. The utility of text mining is demonstrated through description of two promising practices and presentation of two detailed examples. The two promising practices are (1) using text analytics to understand and minimize course withdrawals, and (2) assessing student understanding and depth of learning in science, technology, engineering and mathematics (STEM) (physics). The two detailed examples are (1) refining survey items on the National Survey of Student Engagement (NSSE), and (2) using text to create a learning analytics system at a community college (City University of New York [CUNY]: the Stella and Charles Guttman Community College, or CUNY Guttman). Results of this study include identification of additional item choices for the survey and discovery of a relationship between e-portfolio content and academic performance. Additional examples of text mining in higher education and ethical considerations pertaining to this technology are also discussed.

FRAMING THE ISSUE OF TEXT MINING

Students generate copious amounts of thick, rich data; however, these data are often unexamined because traditional qualitative methodologies used to examine thousands of submissions require extensive resources. Text mining (the machine coding of text with the goal of integrating converted submissions with quantitative methods) offers timely, accurate, and actionable assistance (Zhang & Segall, 2010). This article presents detailed information on how text mining can be used by staff who work in institutional research (IR) and collect, but often are forced to neglect, text-based data.

Examples of text data accessible to IR staff are

- Application essays,
- Written assignments,
- Open-ended survey responses,
- Course Management Software (CMS) postings,
- Student blogs,
- Course evaluations,
- Surveys, and
- E-portfolios.

IR professionals recognize the depth of text data that are—or potentially can be—collected, but might be uncertain how to process those data and use them for campus research informing decision-making. This article presents an overview of the concepts and strategies of text mining and text analytics, and acquaints the reader with the terminology, methodology, and software associated with this technology. Two examples using text data in higher education illustrate how mining can be carried out. It is hoped that the reader will develop a fundamental understanding of text mining, be able to suggest how it can be used to aid decision-makers, and understand the advantages as well as the constraints of this approach to data analysis.

BASIC CONCEPTS

Before describing how the researcher would approach text data, it is important to distinguish between data mining information and analytics. Data mining often implies that the researcher is employing algorithms to explore large data sets (Baker & Yacef, 2009; Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Analytics incorporates data mining techniques to create "actionable intelligence," meaning information that guides decision-making (Campbell, DeBlois, & Oblinger, 2007, p. 42). This information is used to predict case-level behavior that will guide intervention (van Barneveld, Arnold, & Campbell, 2012). In education, learning analytics takes the form of collecting real-time data to measure the effectiveness of teaching practices for a particular student, and to suggest intervention in relative real-time in the case that they are not effective (Suthers & Verbert, 2013). As will be described later, the timing of when data are collected, processed, implemented, and acted on is critical from moving data mining to learning analytics (Arnold & Pistilli, 2012). This article uses similar definitions when describing text mining and text analytics.

A variety of related definitions for text mining can be found in the literature. The following definition is an adaptation of data mining that emphasizes the type of data being mined: "the discovery of useful and previously unknown 'gems' of information from textual document repositories" (Zhang & Segall, 2010, p. 626). Other definitions involve the use of specific methodological/technological approaches: "Text mining is a young interdisciplinary field which draws on information retrieval, data mining, machine learning, statistics and computational linguistics" (Singh & Raghuvanshi, 2012, p. 139).

The rationale for text mining stems from "the need to turn text into numbers so powerful algorithms can be applied to large document databases" (Miner et al., 2012, p. 30). According to Miner et al., text mining and text analytics are broad umbrella terms describing a range of technologies for analyzing and processing semi-structured and unstructured text data. Text mining is the practical application of many techniques of analytical processing in text analytics.

There are several ways that researchers can use software to access, label, and study text data. Familiar to anyone who has used an Internet search engine, information access uses recall systems common in most text-mining approaches; however, the absence of generating new information excludes this practice from true text mining status (Hearst, 1999). Another step toward text mining includes the categorization of text (such as categorizing libraries or academic journals) that can be—in its own right—a means to mine text. Similar techniques can be used in the processes of clustering documents and mining Web content. Beyond pulling specific words or clustering proximity of terms, researchers also intend to extract meaning from the text they study. Those who study computational linguistics contribute to text mining by developing algorithms (a subfield called natural language processing) to measure sentiment and meaning from text. Educational data mining can utilize text in a unique way to measure student learning of important concepts or reflecting on developmental milestones (Baker & Yacef, 2009).

Researchers familiar with qualitative research might be skeptical about the effort of those who work in natural language processing and wonder, "Why wouldn't the researcher simply perform traditional qualitative data methods with text data?" Despite the development of intelligent graders and machine learning, computer programs do not have the capacity to interpret the nuance of writing at the level of the human brain. Nonetheless, there are several reasons why text mining is a viable research technique.

First, data sets can include thousands, millions, or billions of text submissions, precluding the use of traditional qualitative techniques. Second, even if the text data were at a size that could be reviewed by researchers, hiring and training coders increases the resources (time and money) needed to complete this process. For most IR departments, hiring a team of coders might not be practical. When analytics projects are used to identify students who need support, the resulting information can come too late to provide that support if data coding, processing, and interpreting takes too long. Having coders introduces the need to account for inter-rater reliability. Third, these data are computed into numeric values so that they can be combined with other sources of data in statistical models built to predict student behavior. Text mining allows for the sizable and efficient processing of text data.

THE PROCESSES OF MINING TEXT

In their book *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, Miner et al. (2012) present a comprehensive step-by-step guide for text mining. Their text-mining methodology is very similar to other types of research: the researcher has to define the purpose of the project, manage the data (seek, organize, clean, and extract model data), evaluate the results, and, finally, disseminate the results. The aspects of data management (organizing, cleaning, and extracting data) are particular to text mining.

First, the researcher generates a corpus, or a collection of documents or cases in which the desired text exists. During this phase, the researcher removes all nonessential information, such as e-mail addresses or Web links. Second, the researcher either uses an established "stop or include" word list or creates one that filters out words that lack informational value (the, a, his) while highlighting words that do have value. This phase also includes limiting the number of terms by accounting for inflection (plural vs. singular, past vs. present) or the word root (teach for teaching, teacher, teaches). Third, the researcher organizes the terms within cases. This process can be done by using a simple binary notation (1 = term is present) or by noting semifrequent but unique terms (i.e., terms that are featured in some, but not all, of the cases). Another way to organize the terms is by using singular value decomposition (SVD), which reduces the input matrix to a smaller version, representing the variability of each case (Berry & Kogan, 2010; Manning & Schütze, 1999). This latter method is important when working with large text data sets that otherwise could take a long time to process.

Once the researcher has organized the data, she can use several ways to extract important information from these cases (Miner et al., 2012):

- Classification. The researcher creates a dictionary to organize terms based on their definition and hierarchical connection; for example, she might file "classes" under the domain of "education."
- Clustering. The researcher groups terms based on the frequency and pattern of their use, compared to the number of students who use those terms.
- Association. The researcher examines the use of text in connection with some event that is occurring. For example, she might want to compare the positive or negative descriptions of faculty teaching, as reflected in course evaluations before and after final grades are posted.
- Trend analysis. The researcher measures the change of text responses to an identical prompt over time.

SOFTWARE TO CONSIDER

Numerous free and commercial software programs are available for text mining. RapidMiner is an open source analytics program featuring tools that include the analysis of text data. The visual point-and-click nature of the interface allows non–computer programmers to access, clean, and analyze their data. The RapidMiner Web site includes numerous resources, and the user community has posted helpful videos on YouTube. The text extensions include easy-to-use functions such as the ability to group documents based on term frequency. Although the base-level package is free, users might want to purchase the professional package that includes more-advanced options. RapidMiner provides sizable discounts to researchers who use their software, and also provides an extension that allows for the export of data into Tableau. This is an ideal program for IR professionals who want to begin experimenting with text data.

Waikato Environment for Knowledge Analysis (WEKA) is another open source program available for download. WEKA uses a system of algorithms through Java to perform analytic functions; Keyphrase Extraction Algorithm (KEA), on the other hand, is an extension of that project focused on text mining. Both WEKA and KEA were developed at the University of Waikato, New Zealand. WEKA is a premier program for machine learning, a field of computer science that emphasizes the development of programs that can recognize patterns and self-evolve, which is relevant in text mining where programs recognize and adapt to changes in text. KEA has a clean function of recognizing text within a corpus. However, the program does not have a graphical user interface and therefore is most accessible to those with computer science backgrounds. Also, although the WEKA Web site continues to collect publications from researchers using their system, as of this writing the KEA Web site has not been updated for some time and has limited resources.

Most IR professionals are familiar with R as a program, but they might not be fluent in the programming language needed to make it work. R is a premier statistical software, in part because it is free, but also because it has been adopted by a large contingent of statisticians worldwide. There are numerous books and articles on how to use R for text mining; however, between the time it would take staff to become expert in the use of R and to teach themselves how to program the text mining, the software can lose its advantage. Nonetheless, for staff familiar with the program, it can be a great platform to begin to explore and experiment with text data.

The commercial products do distinguish themselves from the free software. IBM SPSS Modeler Premium offers text mining in addition to the analytics suite; the suite includes other data analysis processes including variable selection, various analysis applications, and visualization options. Like most SPSS platforms, the interface is extremely user friendly, probably to a fault, considering how statistical analysis is reduced to a few button clicks. However, for the purposes of text mining, the software is quite advanced. The text function has a built-in dictionary that classifies terms and allows the user to create a dictionary. This function creates a hierarchy of similar terms and places them into broad categories such as "athletics," "emotion," and "education." The application also has built-in dictionaries that include a classification scheme for terms in areas such as customer satisfaction. The natural sorting of terms along with the ease of creating a dictionary allows for comprehensive and easy classification of text data.

SAS® Enterprise Miner™ software also has a user interface and offers comprehensive tools for organizing text and clustering cases.[1] Like IBM SPSS Modeler Premium, SAS Enterprise Miner software is an analytics suite that includes some text-mining applications. However, the way this product distinguishes itself from others is that the clustering function of terms is a component of the text-mining function. Often programs will reduce terms to dichotomous values (not-present vs. present) and then employ clustering methods afterward. SAS Enterprise Miner software allows the user to augment the clustering parameters within the text-mining function and creates data visualizations that are more compelling because of their use of text-mining terms. These features within SAS Enterprise Miner software offer a comprehensive way to cluster terms for analysis.

PROMISING PRACTICES

In higher education there are numerous opportunities to mine text data to predict important student behavior. For example, application essays can be mined to predict student matriculation or even retention. These data can be incorporated into an analytics project aimed at awarding the appropriate amount of financial aid needed to secure a student's enrollment. Another way in which text mining could be used would be to measure virtual student participation in course management software, such as evaluating students' contribution in a course message board within a learning management system like Blackboard or Canvas. Already researchers are using data to model and predict students' academic performance and assign intervention, such as e-mail notification, conversations with advisors, and alerts to faculty (Arnold & Pistilli, 2012). These efforts can be enhanced with the analysis of text data.

Using Text Analytics to Understand and Minimize Course Withdrawals

Two current and promising applications of text analytics involve its application in the areas of student course retention and course outcomes. In mining open-ended comments

[1] SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

captured from students withdrawing from college courses, Michalski (2014) produced and tested a model that succeeded in categorizing over 95% of student explanations for student course withdrawal decisions. This model includes 11 major categories for course withdrawal (i.e., reasons) and corresponds well to existing theoretical and empirical research in the area of college course withdrawal. Of the 11 categories, the top three coding categories were (1) time–schedule, (2) job–work, and (3) personal–other reasons provided by students. Other categories included finances, health, family, course/faculty negative, and online course (mentioned by students who stated their desire to take the same course via instructor-led, rather than online, delivery). Subsequent research (Michalski, 2015) further demonstrated how output from the resulting text model can be statistically analyzed using selected quantitative procedures (including Hierarchical Agglomerative Cluster Analysis, Principal Components Analysis, and Multiple Correspondence Analysis) for validation (i.e., creating clusters of terms describing course withdrawal that are both mathematically and conceptually related). These results are currently being used to develop a course Re-enrollment Assessment Online (REASON) process to minimize course withdrawals, in part by identifying and providing appropriate support, services, advising, and other assistance to encourage and assist student course reenrollment decisions.

Assessing Student Understanding and Depth of Learning in STEM

A second promising example of the application of text mining is the analysis of student responses to open-ended assessment questions at Michigan State University's (2015) Collaborative Research in Education, Assessment and Teaching Environments for the fields of science, technology, engineering and mathematics (CREATE for STEM). There, researchers analyzed student responses to open-ended test questions and related these answers to course outcomes. Within this project at Michigan State, as part of a National Science Foundation grant, Park, Haudek, and Urban-Lurain (2015) used IBM SPSS Modeler to mine the text of short-answer physics test questions about the course topic "energy." The purpose of this study was to explore the degree to which term use is associated with overall knowledge of energy-related constructs. The researchers were able to classify terms used in open-ended responses as either surface-level understanding or scientific ideas. Not surprisingly, students who wrote using scientific ideas were more likely to answer corresponding multiple-choice questions correctly. Innovative uses of text mining like these allow for a more robust understanding of student learning, and assist in test design.

TWO DETAILED EXAMPLES

This article now demonstrates the utility of text mining through presentation of two detailed examples: (1) refining survey items on the NSSE, and (2) relating e-portfolio text to student performance at a community college (CUNY Guttman).

Example One: Classifying Open-Ended Survey Questions

When creating a survey, researchers strive to develop closed-ended items that present all of the possible answers for a given question (Dillman, Christian, & Smyth, 2014). Answering open-ended questions requires more mental energy of the respondent and can lead to answers that are more difficult than closed-ended items for the researcher to process. It can be challenging for institutional researchers to develop a list of all the possible survey responses to a particular question. What do researchers do if they are not sure if all possible responses are presented? A simple answer is to create a text box next to an "other" option where a respondent could write in an answer. Using text mining, the researcher can organize these write-ins to create a more comprehensive list of options.

In the 2014 administration of the NSSE, an experimental item set was developed to identify types of leadership positions held by students. In the development of this survey item set, researchers wanted to know how often students said they had held a formal leadership position in the core survey, as compared to how often they identified serving in a specific

leadership role later in the survey. Of the students who responded to the items, 1,482 of 4,836 (31%) wrote in a leadership position not listed in the core survey item. Text mining was used to determine (1) what new leadership roles should be added to the survey set and (2) the degree to which responses to the leadership item in the core survey related to answers in the experimental item set.
Write-in entries were classified using the text-mining capability of IBM SPSS Modeler Premium. Of the 1,482 students who entered a response, 830 (56%) entries were summarized into eight categories. For example, the concept "teaching assistant" included variations in term, such as "teaching assistant," "teachering assistant," and "teacher's assistant." The top eight leadership positions written into the open-ended text box and their percent of total written comments are shown in Table 1.

The researchers then calculated how well these positions represented the respondents' understanding of a "formal leadership position." In the core survey, respondents were asked if they had completed (or were in the process of completing) a formal leadership position. By comparing the number of students who identified one of these positions to the number of students who said they had participated in a formal leadership position, the researchers had a better understanding of positions that constitute "formal leadership" from the perspective of the respondents. For example, students who wrote in "secretaries," "treasurers," and "editors" were more likely to have said they had completed a leadership role.

Table 1. Counts and Percentage of Write-ins of Additional Leadership Positions Added Through Text Mining

Position	n	%
Tutoring	145	9.8%
Teaching assistant	87	5.9%
Research assistant	60	4.0%
Secretary	55	3.7%
Treasurer	57	3.8%
Mentor	54	3.6%
Member	51	3.4%
Editor	25	1.7%

As a result of this study, the response options "Instructor or Teaching Assistant," "Tutor," and "Editor" were added as leadership positions for the second administration of this experimental item set. Researchers considered installing skip logic that would allow only affirmative responses to the formal leadership item on the core survey to access the experimental item set. However, since a large number of respondents who held positions (both the original and write-in options) had not reported completing a formal leadership experience in the core survey, the skip logic was not included. In this instance, text mining allowed the researchers to expand the item bank and improve the survey design.

Example Two: Clustering E-Portfolio Submissions

As part of a Bill & Melinda Gates Foundation–funded project, researchers with CUNY Guttman were interested in mining first-year e-portfolio introductions to see if student text data could predict academic performance. All students were required to attend a summer bridge program prior to their first semester. During that program, students began their e-portfolio by writing introductions of themselves. Some submissions were informal, reading like an online "shout out" instead of an academic piece of work. Other submissions described ambitions and backgrounds.

Using SAS Enterprise Miner, terms from 163 student e-portfolio introductions were clustered to predict student outcomes such as the number of credit hours earned and GPA. The terms were grouped using a K-cluster analysis within the Enterprise Miner suite; groups of terms were given names by researchers, based on their seemingly shared concept. These concepts, which aligned with student development theory, emerged from the terms; for example, worrying about making friends (e.g., shy, person, friend, know, quiet) and commitment to society (e.g., social, worker, work, believe, help). These results are summarized in Table 2.

Table 2. Results from Text Clustering of Student E-Portfolio Introductions

Concept	Clustered Terms
Connection to family	family, york,* high school, college, child
Learning	class, teacher, art, math, subject
Everyday	know, day, love, life
College participation	high school, school, attend, guttman
Gaming	game, movie, favorite, watch, video
Worrying about making friends	shy, person, friend, know, quiet
Recreation	art, basketball, play, sport, travel
Commitment to society	social, worker, work, believe, help
Technology integration	technology, information, art, health, mind
Aspirations to work in business	guttman, business, manhattan, administration, graduate

Note: CUNY Guttman is located in Manhattan, so it might be common for students who are describing their connection to family and education to include their connection to New York City.*

Variables were coded dichotomously based on whether the term was contained in a concept's cluster. In an ordinary least squares (OLS) regression analysis, after controlling for college preparation (SAT and writing proficiency scores) and age, a relationship ($p = 0.06$) was found between the concept "connection to family" and credit hours earned that fall. Students who used terms within this concept earned fewer credit hours than students who did not. Besides identifying the topics that students write about in their e-portfolio introductions, the results of this research identified one concept—connection to family—that predicted the number of credit hours earned. In this case, students who wrote about connection to family earned fewer credit hours than students who did not. This finding is consistent with the Theory of Student Departure, in which Tinto (1987) argues that family obligations can lead to student attrition. The ultimate goal of this project was to explore whether text mining could be used in learning analytics at CUNY Guttman. A limitation of this study is that this institution has a small enrollment, so many of the principles of big data (specifically a large number of cases or students) will not work. However, by identifying which types of information (such as e-portfolio data) are predictive of student success, researchers at this institution can create analytic models that are accurate, using few variables.

IR professionals often have access to text data such as admissions essays, answers to online tests, and e-portfolio submissions, which are often not incorporated in data analysis because they are too difficult to synthesize compared with basic metrics such as SAT score or GPA. However, clustering terms from these corpuses can help to identify themes within these documents. Although clustering terms requires more mathematical training and the naming of the concepts is subjective (as it is in factor analysis), the mathematical grouping of terms provides insight into what students are writing about, while providing units of analysis that can be included in models to predict student success. Clustering student text provides a means to harness these often-overlooked data.

CONSIDERATIONS FOR IMPLEMENTATION

Text mining has the possibility of being a featured tool in analytics projects; however, institutional stakeholders need to be intentional in how they collect and process these data so they can use information in a timely manner. Processes for data collection, distribution, project implementation, and analysis of results must be coordinated. In any analytic project, timing is important. Text data need to be cleaned, processed, and incorporated into predictive models in time for interventions to occur; because of the quick turnaround, most projects include automation.

Another aspect to consider is the data source or origin. The two examples in this article illustrate different sources of text data. In the first example, asking respondents to submit one "other" leadership position seems straightforward and easy to sort; however, there can be issues with classifying even these basic data. For example, some responses do not easily or naturally fit into a category, while others might be placed in more than one category. In the second example, researchers were mining students' introduction statements for specific information, but the prompts might have been too vague. Alternatively, prompts can be too prescriptive. Contrast the prompts, "What do you think it takes to succeed in college?" with "Who will you ask to mentor you so you will be successful in college?" Similar to survey item design, the creation of prompts for text mining might develop into its own science. (For information regarding the intersection of text mining and survey question design, see George et al., 2012.)

Another aspect to consider when using text data for predictive models is the use of sentiment analysis: "Is it only important to know what a student is writing or is it also important to understand how a student feels about a particular topic?" There might be text sources, such as student course evaluations or satisfaction surveys, where it would be important to consider sentiment. In these cases, the sentiment analysis component of customer satisfaction software can be implemented to measure student attitude when describing certain phenomena as positive, negative, or indifferent. IBM SPSS Modeler Premium has a built-in library for measuring customer satisfaction, allowing for a more refined and adjustable view of individual likes and dislikes. Measuring sentiment might be crucial for stakeholders (e.g., administrators and faculty); therefore, for researchers implementing text-mining projects, the analysis of sentiment might be an important aspect when trying to garner buy-in to approve text-mining projects.

ETHICAL CONSIDERATIONS

As we continue to evolve in this era of analytics, it is not uncommon to hear concerns about how data are being collected and used. Students might believe that their intellectual property or even identity has been exploited when researchers use their text data, some of which can be very personal (Slade & Prinsloo, 2013). Because there is much at stake in terms of student response to analytics, institutions using these technologies need to develop sophisticated policies regarding the data ownership, transparency, and security of data (Pardo & Siemens, 2014). Furthermore, institutions are sometimes ill-equipped to grapple with the more complex issues around the use of data within analytics, such as what to do when unexpected outcomes occur, such as predictive models that misclassify student potential (Willis, 2014). All of these factors reflect a campus culture in which administrators act ethically in using data to improve student success, to ensure that students and faculty are valued partners with technologies, and to create an environment that views these technologies as advancing those aims (Willis, Campbell, & Pistilli, 2013). These concerns over analytics also relate to the use of text mining.

One of the benefits of text mining is that it presents a bridge between the atheoretical process of data mining/analytics and the more theoretically driven approaches for research used in higher education. It can be unnerving for campus stakeholders, including faculty, administrators, and students, to rely on the black box of analytics. Text mining raises ethical concerns about the results of analytics processes. What is the implication when an algorithm predicts that a student is less likely to pass a course or persist within a particular major? What aspects of a student's background and institutional environment are either being overlooked or considered in making these predictions? Furthermore, institutional policy is traditionally built on empirical studies of students that reference some theoretical framework, or at least a plausible hypothesis. For example, first-generation college students are disadvantaged, and therefore administrators argue that they should receive additional resources as a matter of policy. Gathering stakeholders to support this one aspect of identity is easy because it is an understandable concept: "Students who did not grow up in a home where at least one parent completed college are at a disadvantage compared with students who did." However, in an analytic project, first-generation status might be one variable among many to predict student success. These complex models replace an understandable narrative such as the one related to first-generation status with a narrative that is reliant on numerous factors. On the other hand, the results of text mining can be easily interpreted. In the second example of this study, there was an inverse relationship between a student mentioning connection to family and earning credit hours. An educator meeting with students and reviewing e-portfolio introductions would find it easy to act on this information.

There are other institutional cultural aspects to consider when implementing text mining programs on campus: Students could react to the common knowledge that all of their submissions on the campus management system are being mined. Students could change their responses once they realize that their text is being mined. In the second example of this study, it might be the case that upper-class students will warn incoming first-years, "Make sure not to mention connection to your family in your e-portfolio introduction; otherwise, you'll have to schedule an additional meeting with your advisor."

When considering equity, institutional stakeholders need to take the increasingly heterogeneous nature of student demography into consideration. Researchers need to account for diverse levels of familiarity with English, resources, and prior education when working with student text data. Simply relying on text data without regard to student background variables can mislead. Students who have better educational preparation or more resources may be advantaged in the types of phenomena they describe in text. If so, text mining could be used to reinforce inequity within higher education. These aspects, though not the focus of this article, ask whether the researcher should mine text data. Researchers need to consider that question at length before they consider whether they are able to mine text data.

PUTTING TEXT MINING TO USE

This article describes the ways in which text mining can be implemented in the work of those institutional researchers who are asked to create analytics programs on their campuses. The article described two promising application areas, and detailed two examples of text mining projects. Important considerations for any text-mining project are the context and timing of text collection. Although text mining offers several promising avenues within higher education, ethical considerations should provoke deep questions about the appropriateness of these methods balanced with the needs of the institution.

Like many projects within an IR shop, successful text mining requires multiple areas of expertise (Haight, 2014). First, expertise in data acquisition and management is needed to procure and then clean data (e.g., to remove extraneous text) while pairing data with other sources of student information. Second, someone must mine and model text data into predictive schemes that are statistically sound. Third, expertise in graphic design or data visualization is needed to successfully communicate the connection between data source, usable text, and student outcomes. Fourth, as described earlier, the use of automation (i.e., creating systems to automatically collect, clean, and process data) is paramount in any analytics process, so a team member needs to be in charge of this aspect of the project. Finally, text-mining projects need a team champion to advertise and demonstrate the value of this technology and to promote it to campus stakeholders. These skills are essential to integrating text mining into campus decision-making. Text mining presents a microcosm of other IR processes while offering an exciting way to make use of dense text data waiting to be unlocked. As a field, IR is in a unique position because text data have been collected for decades and the complex decision-making on college campuses can only be further informed by the inclusion of those data.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). doi:10.1145/2330601.2330666.

Baker, R.S.J.D., Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.

Berry, M. W., & Kogan, J. (Eds.). (2010). *Text mining: Applications and theory*. Hoboken, NJ: Wiley & Sons.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42(4), 40–57.

Dillman, D. A., Christian, L. M., & Smyth, J. D. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley & Sons.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.

George, C. P., Wang, D. Z., Wilson, J. N., Epstein, L. M., Garland, P., & Suh, A. (2012, December). A machine learning based topic exploration and categorization on surveys. In *11th International Conference on Machine Learning and Applications* (vol. 2, pp. 7–12). doi:10.1109/ICMLA.2012.132

Haight, D. (2014). *The five faces of analytics*. Dark Horse Analytics. http://www.darkhorseanalytics.com/blog/the-five-faces-of-analytics

Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3–10). doi:10.3115/1034678.1034679

Manning, C. D., & Schütze, H. (Ed. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Michalski, G. V. (2014). In their own words: A text analytics investigation of college course attrition. *Community College Journal of Research and Practice*, 38(9), 811–826.

Michalski, G. V. (2015). *Using analytics to minimize student course withdrawals.* Paper presented at the Association for Institutional Research (AIR) Annual Forum, Denver, CO, August. http://www.forum.airweb.org/2015/Documents/Presentations/1095_6bd9881f-0d68-457f-8557-8da3d450cbb8.pdf

Michigan State University. (2015) *Home page*. CREATE for STEM Institute. http://create4stem.msu.edu/

Miner, M., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, B. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Waltham, MA: Academic Press.

National Survey of Student Engagement (NSSE). 2014. NSSE 2014 experimental items codebook formal leadership. http://nsse.indiana.edu/pdf/exp_items/2014/NSSE%202014%20Exp_FOL_Codebook.pdf

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.

Park, M., Haudek, K., & Urban-Lurain, M. (2015). Computerized lexical analysis of students' written responses for diagnosing conceptual understanding of energy. In *National Association for Research in Science Teaching (NARST) 2015 Annual International Conference*, April. http://create4stem.msu.edu/publication/3361

Singh, P. D., & Raghuvanshi, J. (2012). Rising of text mining technique: An unforeseen-part of data mining. *International Journal of Advanced Research in Computer Science and Electronics Engineering*, 1(3), 139–144.

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.

Suthers, D., & Verbert, K. (2013). Learning analytics as a middle space. In Proceedings of the *Third International Conference on Learning Analytics and Knowledge* (pp. 1–4). Leuven, Belgium, April 8–12.

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

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative, 1,* 1–11.

Willis III, J. E. (2014). *Learning analytics and ethics: A framework beyond utilitarianism*. EDUCAUSE Review. http://er.educause.edu/articles/2014/8/learning-analytics-and-ethics-a-framework-beyond-utilitarianism

Willis, J. E., Campbell, J. P., & Pistilli, M. D. (2013). *Ethics, big data, and analytics: A model for application*. EDUCAUSE Review. http://er.educause.edu/articles/2013/5/ethics-big-data-and-analytics-a-model-for-application

Zhang & Segall, 2010 Zhang, Q., & Segall, R. S. (2010). Review of data, text and web mining software. *Kybernetes*, 39(4), 625–655.

Thank You!

AIR expresses sincere appreciation for the members who serve as peer reviewers. Authors who submit materials to AIR Publications receive feedback from AIR members. The following individuals reviewed manuscripts submitted to this volume of *Professional Files*.

Rebecca Barber

Gabriela Borcoman

Maggie Dalrymple

Chris Davis

Lauren Friedman

Fernando Furquim

Gina Johnson

JJ Juszkiewicz

Wendy Kallina

Kathi Ketcheson

Amy Kline

Ebenezer Kolajo

JoLanna Kord

Barry Nagle

Amy Prettejohn

Liz Sanders

Justin Shepherd

Joe Viscomi

Newman Wong

Julie Wren

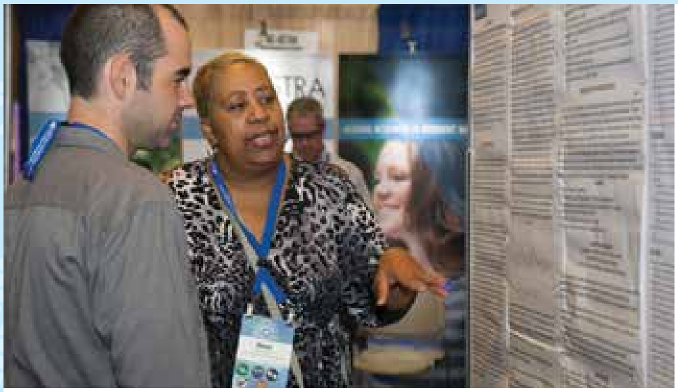
Hongtao Yue

About AIR *Professional File*

Professional File volumes are journal-length publications grounded in relevant literature that synthesize current issues, present new processes or models, or share practical applications related to institutional research. All submissions are peer-reviewed. For more information about AIR Publications, including *Professional File* and instructions for authors, visit www.airweb.org/publications.

Association for Institutional Research
Jason R. Lewis, Interim Executive Director
www.airweb.org

ISSN 2155-7535



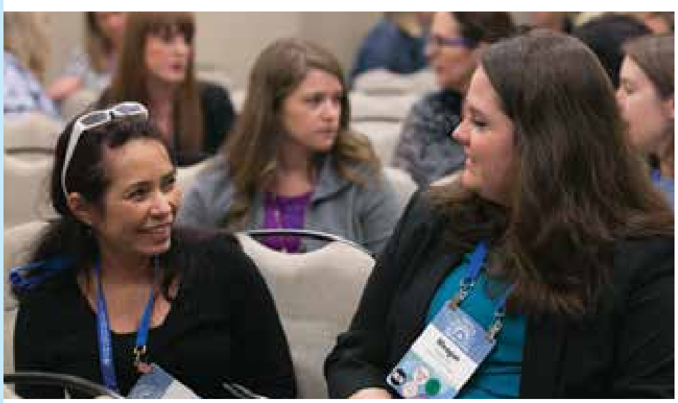
IT ALL STARTS WITH A CONVERSATION...

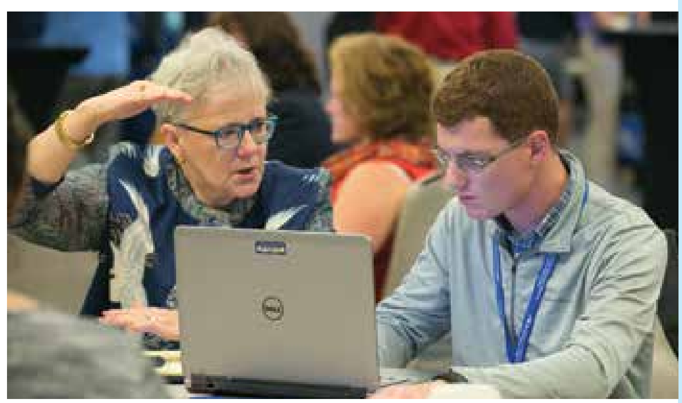
AIR's educational resources, best practices, and professional development opportunities are designed to help professionals in IR and related fields effectively use institutional research – data, information, and analysis for decision support – in a rapidly changing environment.

» **Digital Pass** – Join your colleagues in discussing the 40+ available recorded sessions from the 2016 Forum.

» **AIR Forum** – Learn, connect, and share at the largest gathering of IR professionals in the world. Keep the conversation going by using our database of presentation slides, handouts, and scholarly papers.

» **A Statement of Aspirational Practice for IR** – Have a conversation with your president or provost about the new vision of effective IR in support of student success.





AIR Association for Institutional Research
Data and Decisions for Higher Education

LEARN MORE AT WWW. AIRWEB.ORG