

**ISSUES IN WEB SURVEYS OF STUDENT POPULATIONS:  
RESPONSE RATES AND POST-STRATIFICATION WEIGHTING**

Joel David Bloom, Ph.D.

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

Research Assistant Professor  
Department of Political Science

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

jbloom@albany.edu  
(518) 437-4791

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

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

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

### **What Works and What Doesn't**

#### Introduction: Why Response Rates Matter in Student Surveys

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

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

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

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

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

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

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

### **The Research Questions**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

### Experiment 3: The 2007 UAlbany Cable Survey

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

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

The results of this experiment are shown in Table 2, below. When the survey was closed on Monday, May 7<sup>th</sup>, 326 of the students who were addressed personally responded (for a 9.4% response rate), compared to only 265 of those addressed generically (for a 7.7% response rate). This translates to a 23% increase in the raw numbers of responses. In addition to being substantively large and in the expected direction, the difference was statistically significant, with a t-ratio of 2.628 ( $p=0.009$ ). Again, the results are in line with other research showing improved response rates associated with personalized salutations.

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

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

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



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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## Summary and Conclusion

The main findings of the six experiments presented here include:

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

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

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

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

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

---

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

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

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

#### Introduction

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

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

In this manner, differential nonresponse may introduce bias in population estimates. (Mandell, 1974)

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

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

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

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

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

### Data and Analysis

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

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

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

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

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

---

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

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

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

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

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

The 2006 Student Opinion Survey

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

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

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



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

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

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

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

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

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

---

<sup>5</sup> For purposes of this paper, I use the SUNY system’s names for racial and ethnic categories, simply because those are the categories that exist in our student data records.

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

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

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

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

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

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

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

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

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

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

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

## The 2007 Student Experience Survey

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

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

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

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

---

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

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

	<b>Frequency</b>	<b>Sample Percent</b>	<b>Population Percent</b>	<b>Difference</b>	<b>Weight</b>
<b>Female</b>					
White	771	38.1%	28.5%	9.6%	0.75
Black	103	5.1%	5.4%	-0.3%	1.05
Hispanic	82	4.1%	4.1%	0.0%	1.00
Asian or Pacific Islander	78	3.9%	2.7%	1.2%	0.70
Amer. Indian or Alaska Nat.	3	0.1%	0.1%	0.0%	0.78
Non-Resident	24	1.2%	0.9%	0.3%	0.72
Unknown	212	10.5%	7.8%	2.7%	0.74
Total, Female	1273	62.9%	49.4%	13.6%	0.78
<b>Male</b>					
White	495	24.5%	30.9%	-6.4%	1.26
Black	23	1.1%	3.3%	-2.2%	2.94
Hispanic	41	2.0%	3.5%	-1.4%	1.71
Asian or Pacific Islander	46	2.3%	2.6%	-0.3%	1.15
Amer. Indian or Alaska Nat.	0	0.0%	0.1%	-0.1%	NA
Non-Resident	7	0.3%	1.0%	-0.6%	1.37
Unknown	138	6.8%	9.3%	-2.4%	1.36
Total, Male	750	37.1%	50.6%	-13.6%	1.37
Total, Sample	2023	100.0%	100.0%	0.0%	
<b>Student Level</b>	<b>Frequency</b>	<b>Sample Percent</b>	<b>Population Percent</b>	<b>Difference</b>	<b>Weight</b>
Freshman	365	18.0	17.1	0.9	0.95
Sophomore	497	24.6	24.9	-0.3	1.00
Junior	525	26.0	28.2	-2.2	1.09
Senior	636	31.4	29.9	1.6	0.95
Total	2023	100.0	100.0		
<b>Admission Type</b>	<b>Frequency</b>	<b>Sample Percent</b>	<b>Population Percent</b>	<b>Difference</b>	<b>Weight</b>
Freshman	1431	70.7	65.0	5.7	0.92
Transfer	590	29.2	35.0	-5.8	1.20
Total	2021	100.0	100.0		

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

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

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

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

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

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

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

Mode of Administration Effect? A Perverse Exercise



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

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

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

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

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

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

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

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

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

The 2009 Student Opinion Survey

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

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

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

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

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

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

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

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

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

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

Conclusion: Why Weight?

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

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

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

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

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

- 1) Stratified Samples. The 2006 SOS sample design was essentially a stratified sample, in which the administrators got as close as they could to producing a probability sample. But the classes chosen were not a true probability sample; as we have seen, some groups had greater and some had lesser probabilities of selection. For this type of sampling methodology, weighting for probability of selection is indeed required in order to conduct any statistical tests on the data. In this case, I in effect combined weighting by differential probability of selection with weighting by nonresponse (see Groves et al., 2004, pp. 323-326).
- 2) You Never Know. Just because one survey didn’t change after weighting doesn’t mean that the next one will not. Thus, it is always worth the small amount of time it takes to try at least a quick first-stage weighting scheme to see if anything jumps out at you. If it does, you can put in the additional time and effort to really do it right; if there isn’t anything there, you can tell people that you checked.
- 3) Not All Items are the Same. Just because some survey items don’t change after weighting doesn’t mean you can be sure that none will. For example, some survey items may be particularly sensitive to student level; others might be more sensitive to race; still others might be more sensitive to gender. We should always keep that in mind when thinking about weighting, and make sure we include relevant weighting variables whenever possible.
- 4) Campus Politics. Suppose that your campus has an undergraduate population that is 10% African American and you issue a survey report showing that in your survey, only 5% of your sample is African American. Some people might not be happy with that, and they would have a point!
- 5) Do the Right Thing. Finally, even if none of these other factors applied, we should still consider weighting whenever we have time to do so. We have the sample demographics; we have the population demographics; if they differ systematically, weighting is simply the right thing to do. We don’t necessarily have to report the weighted results (especially when they show only minor differences), but even in these cases, weighting is still valuable insofar as it increases our confidence in the validity and reliability of our survey results.

## REFERENCES

- Bloom, Joel (2008). Worth the Weight?: The Benefits and Pitfalls of Applying Post-Stratification Weights to Web Surveys of College Undergraduates. *Presented at the 2008 Annual Meeting of the American Association of Public Opinion Research, New Orleans, Louisiana.*
- Couper, Mick (2008). *Designing Effective Web Surveys*. Cambridge: Cambridge University Press.
- Crawford, S.D., McCagbe, S.E., Saltz, B., Boyd, C.J., Freisthler, B., and Paschall, M.J. (2004). "Gaining Respondent Cooperation in College Web-Based Alcohol Surveys: Findings from Experiments at Two Universities." Paper presented at the 2004 annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona.
- Dillman, Don A., Smyth, Jolene D., and Christian, Leah Melani (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. New York: Wiley.
- Groves, Robert M., Floyd J Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau (2004). *Survey Methodology*. New York: Wiley.
- Heerwegh, D. (2005). Effects of Personal Salutations in E-Mail Invitations to Participate in a Web Survey. *Public Opinion Quarterly*, 69:588-598.
- Joinson, A.N. (2005). "Audience Power, Personalized Salutation and Responses to Web Surveys." Paper presented at the ESF Workshop on Internet Survey Methodology, Dubrovnik, Croatia.
- Joinson, A.N., and Reips, U.-D. (2007). Personalized Salutation, Power of Sender and Response Rates to Web-Based Surveys. *Computers in Human Behavior*, 23:1372-1383.
- Joinson, A.N., Woodley, A., and Reips, U.-D. (2007). Personalization, Authentication and Self-Disclosure in Self-Administered Internet Surveys. *Computers in Human Behavior*, 23:275-285.
- Kaplowitz, Md., Hadlock, T.D., and Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, 68:94-101.
- Mandell, Lewis. (1974). When to Weight: Determining Nonresponse Bias in Survey Data. *Public Opinion Quarterly*, 38:247-252.
- Pearson, J., and Levine, R.A. (2003). Salutations and Response Rates to Online Surveys. Paper presented at the Association for Survey Computing Fourth International Conference on the Impact of Technology on the Survey Process, Warwick, England.
- Porter, S.R., and Whitcomb, M.E. (2003). The Impact of Lottery Incentives on Student Survey Response Rates. *Research in Higher Education*, 44:389-407.
- Singer, Eleanor, ed. (2006). *Special Issue: Nonresponse Bias in Household Surveys*. *Public Opinion Quarterly* 70:5.
- Trouteaud, A.R. (2004). How you ask counts: A test of internet-related components of response rates to a web-based survey. *Social Science Computer Review*, 22:385-392.