Paper-and-Pencil versus Web Administration of a Student Satisfaction Survey

Chieh-Chen Bowen
Cleveland State University
Department of Psychology
2121 Euclid Avenue
Cleveland, OH 44115, USA.

Abstract

Web surveys have been proven more cost efficient and less time-consuming than the traditional paper-and-pencil method. This study demonstrated comparable results from different modes of administration of the same survey under conditions where the intended population remained the same, and equal access to internet was established for every member in the targeted population. A student satisfaction survey administered via the web generated comparable results as the paper-and-pencil method in terms of scale reliability, average scale ratings and inter-correlations among scales. Results supported that web surveys could be used as a cheaper and faster alternative of paper-and-pencil surveys in university settings.

Key words: Survey methodology, Survey modes, Web survey, Paper-and-pencil survey, Measurement equivalence

Introduction

Researchers have taken advantage of the great benefits associated with the accessibility of the internet and its applications specifically to survey administration. Although surveys are widely accepted as a useful research tool to obtain information regarding people's attitudes, preferences, and opinions, the methodological transition from traditional paper-and-pencil surveys to web surveys have provoked many empirical questions of comparability. Differences in samples attributable to different methods, response rates, demographic composition, and the nature/effects of self-disclosure are just a few of the areas that have been widely discussed (Schmidt, 1997, Smith & Leigh, 1997) but the differences of the survey modes were inconsistent in either directions. Until recently, multiple research projects compared web and in-person survey modes and found minimal statistically significant differences in responses to questions about risky behaviors among web surveys, paper-and-pencil surveys and in-person interviews (Eaton, Brener, Kann, Denniston, McManus, Kyle, Roberts, Flint, and Ross, 2010; McMorris, Petrie, Catalano, Fleming, Haggerty and Abbott, 2009). Many attitude surveys also showed that regardless of the medium, the same attributes appeared to be assessed with little distortion (Booth-Kewley, Edwards, & Rosenfeld, 1992; Stanton, 1998). So far, mixed-mode surveys have found web surveys to be a promising alternative to the traditional paper-and-pencil surveys.

While many benefits, potential problems and solutions of using world-wide-web that has the potential of reaching every person on earth have been addressed by Schmidt (1997), this study does not have the ambition to be far-reaching. The purpose of this current study was to empirically investigate measurement equivalence between web surveys and paper-and-pencil surveys when both surveys were designed to reach the same target population and the internet access was provided for every member in the population. University campus is an ideal setting to test measurement equivalence of results generated by different survey modes based on its capability to provide internet access to every student, staff and faculty and its internet security measures to limit access only to its constituents.

Administering a survey via the intranet on a university campus to the constituency served by the university seems to be highly desirable due to the fact that it brings a wide range of the benefits and provides solutions to potential problems. First, the intended population stays the same either for the web survey or paper-and-pencil survey when conducting a student satisfaction survey on a university campus. The intended population is the student body. Second, equal access to the internet is guaranteed for every member in the population. Surveys can be accessed in a controlled fashion (i.e. using student ID to log in the university's intranet), and the constituents are mostly computer literate and all have access to the university's intranet.

Any student who wants to use the intranet can get connected on a university campus. Therefore, equal access to the intranet is ensured for the entire student body. Universities can be viewed as institutions with a dynamic interaction of faculty, students, and administrators. Increased communications among these groups can only lead to a more effective system for engaged learning. Student satisfaction is a construct that is of great interest to the university administration. According to Corts, Lounsbury, Saudargas, and Tatum (2000) increased student satisfaction can contribute to student retention and academic, personal, and professional achievement. An accurate assessment of student perceptions of the quality of service on campus is an important step to help a university to identify and prioritize crucially weak service areas and justifiably spend time, budget and effort to improve them. In addition, positive results from a student satisfaction survey serve as reinforcements for those areas where a university is performing well.

A student satisfaction survey needs to be done in a manner that produces unbiased and representative results. Unbiased and representative results can be achieved by obtaining a randomly selected large sample. A large random sample usually requires a considerable amount of time and money which may be infeasible, impractical or cost-prohibitive. The purpose of this study is to explore using a large sample of a traditional paper-and-pencil survey as the anchor and compare the results of a small and completely voluntary web sample to examine the comparability of the results. The time-saving and cost efficient benefits of a web survey only make sense if the accuracy and representativeness of the results are not jeopardized.

A satisfaction survey specifically designed to quickly gauge students' perceptions on the quality of services they receive on campus is a specific application of an opinion or attitude survey. There is no right or wrong answer to such a survey. It is designed to provide information to describe and understand the current status of the topics in the survey. Such a survey can be administered as either a web survey or a paper-and-pencil survey or both. Universities are eager to adopt web administration of student satisfaction surveys to save time and money in collecting data as well as setting surveys up as an automatic annual process to compare results longitudinally. However, whether a web survey can produce unbiased and representative results equivalent to those in a large sample remains an empirical question.

Eaton, et al. (2010) conducted a web survey and a paper-and-pencil survey in educational settings across 85 schools and 15 states asking high school students about risky behaviors and found 7 out of 74 risky behaviors showed significant differences associated with survey modes. Although the results demonstrated measurement equivalence between two survey modes and produced reasonable estimates of the risk behaviors, the researchers cautioned against switching the traditional paper-and-pencil survey to a web survey without first implementing security measures to protect students' privacy.

Cronk & West (2002) investigated a web versus paper-and pencil administration of a personality questionnaire and found that results showed no significant mean scores differences using the Internet or paper-and-pencil methods. This trend is consistent with other previous studies (Buchanan & Smith, 1999; Davis, 1999; Stanton, 1998). Results of paper-and-pencil questionnaires on one hand and web-based questionnaires on the other indicated similar internal consistency and factor structures. Researchers may accordingly proceed with a guarded assumption of measurement equivalence between paper-and-pencil and web administration of the same instrument.

Many empirical studies have demonstrated the comparability of results between a computerized version and a paper-and-pencil version of the same instrument. Stanton (1998) compared a web survey and a paper survey on supervisory fairness and showed identical factor structures to support equal quality of the data collected from different modes. Also the similarity in inter-scale corrections between the two samples supported the exploration of construct equivalence. Vispoel (2000) investigated 212 college students' self-concept using a within-subject design. Results of that study showed strong support for comparable results between computerized versus paper-and-pencil versions of the self-concept assessment from the same person. First, it was found that regardless of presentation order (computerized assessment first or paper-and-pencil first) the computerized version took longer to complete with a 102.3 second difference. Second, a multiple analysis of variance revealed that there were no significant differences in variability with respect to the mode of administration. The modes did not differ with respect to reliability or validity. Similar rank ordering of the respondents in both modes suggests that the results were comparable.

Buchanan and Smith (1999) administered a personality questionnaire, the Revised Self-Monitoring Questionnaire, to 936 web respondents and 224 students in classrooms. Both samples generated fairly similar reliabilities between 0.73-0.75. A confirmatory factor analysis was conducted on the two samples to compare with previously reported factors commonly associated with the instrument. Results indicated that both modes had respectable levels of goodness of fit, with similar levels of internal consistency and factor structures, and that these modes of questionnaire administration were comparable.

Previous studies have established comparability of results between web-based administration versus paper-and-pencil administration using personality inventories in both a within-subject design (Vispoel, 2000) and a between-subject design (Buchanan & Smith, 1999). Opinion surveys are similar to personality inventories in that they both involve self-reported responses and neither has 'right' or 'wrong' answers to the questions. Web surveys have been commonly thought to generate potential biased results because only people with specific characteristics (i.e. highly educated, computer savvy, or with strong opinions about the topic) would self-select to respond to web surveys (Stanton, 1998). However, it remains untested whether a web-based opinion/attitude survey can produce unbiased and representative results as in a large sample using paper-and-pencil method when the target population remains the same for both survey modes.

Tomsic, Hendel, and Matross (2000) conducted a Student Experience Survey using internet and paper-and-pencil formats over a three-year period (1997-1999). During the three-year of data collection, significant differences between the results from the two modes appeared at first, and diminished at the end. This diminution of difference may have been due either to the two groups becoming more similar or due to the improvements at the university. The authors concluded that web surveys might serve as the predominant format for surveys of students, faculty, and staff.

In the past, traditional paper-and-pencil surveys have done an excellent job of facilitating student feedback. Due to the fact that universities are able to supply Internet connection to all their staff, faculty and students, University administrations may possibly benefit from a web-based method of assessing students' opinions. However, before this mode can be used confidently as the sole method for collecting students' opinions, it is necessary to empirically establish that web-based surveys can be as effective as traditional paper-and-pencil surveys at accurately soliciting and describing students' opinions.

Specifically, the current research investigates comparability of the two modes of survey administration (paper-and-pencil vs. web) regarding specific issues such as sample similarity, and response equivalence, and inter-correlations of the ratings when the paper-and-pencil sample is designed to be representative of the population. Accordingly, I investigated the following three research questions.

> *Question 1: Is a large paper-and-pencil sample representative of the entire student body at the University in terms of the annually reported demographic information such as gender, race, colleges and number of credit hours which they are currently enrolled?.*
> *Question 2: Is there measurement equivalence between the paper-and-pencil and web modes of the same survey?*
> *Question 3: Are results of the web survey similar to the results of the large paper-and-pencil sample?*

Method

Participants

For a sample to be representative of the population, random selection in which every member in the population had equal chance of being selected is the ultimate ideal. However, this ideal is usually not realistic or feasible due to time and other resource constraints. To make the sampling process more manageable in reality, in this study, students were selected based on the list of classes offered in the spring semester to participate in the paper-and-pencil survey. A single-stage cluster sampling was conducted where random sampling occurred at the class level and every student within the class was included in the survey. Cluster sampling is appropriate when natural grouping appeared in the population. In this case, within the university setting, classes were natural grouping. The main objective of cluster sampling is to reduce costs by increasing sampling efficiency (Shadish, W. R., Cook, Thomas D., & Campbell, D. T. (2002).

Approximately thirty percent of the classes throughout the university were randomly selected and a letter requesting access to the class was sent out to the instructor on record. One hundred and fourteen instructors (38% of those contacted) allowed researchers access to their classes. Such a response rate is in line with an average 35% response rate for the National Survey for Student Engagement (NSSE) paper administration across hundreds of college and university campuses in 2005. A total of 2282 students completed the paper-and-pencil survey in class. The population of the students in the university was 15,998 at the time of the survey.

The appropriateness of a survey sample largely depends on the purpose of the survey. If the content of the survey were class-specific, for example, questions regarding satisfaction on the instructor's presentation skills or instructor's knowledge, the answers were expected to be more similar within each class but varied across classes. However, the content of this survey focused on perceived quality of services provided on campus, including Registrar's office, Bursar's office, admissions, advising, parking, bookstore, and food services. Enrolling in different classes is not likely to have systematic impact on the perceived quality of services on campus.

Besides the paper-and-pencil survey, a Web survey was set up simultaneously during the same three-week period data collection. The survey link was posted on the home page of the university website and participation was completely voluntary. An article in the university campus newspaper informed students of the survey's purpose and requested students' participation. At the same time, a mass mailing was sent to all registered student's electronic mailing address. Students were provided with the necessary information, including a link to the web address of the survey. Randomly drawn prizes were used to encourage participation. Prizes included a free parking pass for a semester, a book scholarship for a semester, $25 cash cards, and opportunities to have dinner with the University President. All participants were required to have a valid student identification number to begin the survey. Two hundred and thirty-seven students had completed the web survey by the end of three weeks.

Materials

The Quality of Service Survey was designed for this particular study. It was created based on the discussions of ten focus groups and an open-ended questionnaire from students. In addition, one focus group was held with each director of the different service areas on campus. Items resulting from these group meetings were piloted on 150 students. Results of the pilot test were used to modify and improve the item clarity and quality of those included in the Quality of Service Survey.

The paper-and-pencil survey included a cover page with instructions to use the rating scale, a warning for the presence of negatively phrased items, and explanations for asking participants to provide academic and demographic information for statistical purposes. The survey consisted of one hundred and six items focusing on various service areas on campus. Fifteen questions were related to the Bursar office, twelve related to the Financial Aid Office, nine related to the Parking Office, fourteen related to the Registrar, nine related to Food Services, eight related to the Bookstore, eight related to Advising, and nine related to the Admissions Office. The remaining twenty-two questions were related to career services, safety, interoffice communication, signage, and the overall university experience. To make the results more concise, individual service areas with three items or less were not included in the results.

The web survey was created to be an electronic page-turner of the paper-and-pencil survey. The order of the items was the same as the paper-and-pencil measure. Respondents were allowed to go back to previous pages as they did with the paper-and-pencil version before submitting the survey. Once they clicked the finish button, they would no longer have access to the survey. However, in the paper-and-pencil version students were asked to fill in a bubble sheet numbered from 1 to 5 (strongly disagree to strongly agree) or 9 (not applicable), while in the web-version participants were asked to select one of these same aforementioned ratings from a "drop-down" menu.

Procedure

When distributing the paper-and-pencil version, all survey administrators provided the participants with a standardized instructions and answers to frequently asked questions. Student identification numbers were used to identify prize winners only and the survey results remained anonymous. The web version of the survey required a student identification number to begin the survey, which allowed verification of enrollment status. An answer was required for each item before respondents were allowed to move to the next. Multiple submissions from any student were addressed by randomly deleting extra submissions.

However, this procedure was only feasible when students provided their student number multiple times. The chance of students who participated multiple times without providing their student numbers existed but was negligible.

Results

Sample representativeness

Sample representativeness was measured by comparing sample demographics with known population demographics. The demographic representativeness of a sample was investigated by comparing its characteristics with population characteristics. A brief list of characteristics of the student population, the paper-and-pencil sample and the web sample was shown in Table 1. Annually reported and most salient demographics of student population were gender, race, colleges and number of credit hours enrolled. To answer the first research question, '*Is the large paper-and-pencil sample or the web sample representative of the entire student body at the University in terms of their gender, race composition, colleges and number of credit hours which they are currently enrolled?.*' Chi-square tests were conducted to compare four demographic characteristics (gender, race, colleges and number of credit hours enrolled) between the two samples and the university student population. The Chi-square statistics as reported in Table 1 revealed that the composition of the large paper-and-pencil sample and the web sample were both similar to the university population in terms of gender but different in terms of race, colleges and number of credit hours enrolled. Therefore, I concluded that the paper-and pencil and web sample reached different audience and were not demographically representative of the entire student body in the university. Even different modes of survey administrations reached different audience, it is still important to exam whether measures were obtained with little distortion.

Measurement equivalence between different modes of survey administration

Measurement equivalence was measured by the comparability of reliability coefficients. The compatibility of reliability between the web and paper-and-pencil survey is documented in Table 2. The reliability coefficients in web sample were highly similar to their counterparts in the paper-and-pencil survey. Significance tests on the difference of paired reliability coefficients were conducted by first transforming the reliability coefficients to Fisher's z_r, then testing the difference between each pair of reliability coefficients in paper-and-pencil and web samples. None of the differences of the Cronbach's α coefficients between web and paper-and-pencil scales were statistically significant (all comparisons yielding, $p > .05$). Both web and paper-and-pencil surveys resulted in almost identical Cronbach's α in all scales. The same Cronbach's α in the paper-and-pencil survey can be found in the web survey, it reflected that whether survey was administered via paper-and-pencil or web had very little impact on the way respondents answered the questions. All scales except *Parking* ($\alpha = .54$ in the web survey and $\alpha = .49$ in the paper-and-pencil survey) reached acceptable reliability at least in one sample. Therefore, the *Parking* scale was excluded from further analyses. The Bookstore scale reached reliability .74 in the paper-and-pencil sample but only .64 in the web sample. Due to the exploratory nature of this survey, the Bookstore scale was retained for the rest of the analyses.

Comparison on scale ratings between web and paper-and-pencil surveys

So far, it was clear that different modes of surveys reached different audience. However, different audience perceived and understood the survey items in a very similar manner. The next question needs to be answered is, '*Are results of the web survey similar to the results of the randomly selected large sample?*' comparisons of the scales means and the comparability of the inter-scale correlations from both samples were used. MANOVA was conducted to test the overall quality of service ratings between web and paper-and-pencil samples. Pillai's Trace for the MANOVA was .023 with F value 3.69 and $p < .000$.

The Bonferroni approach was used to adjust the analysis-wide type I error at .05 when conducting a series of F-tests. Since eight F-tests were conducted simultaneously, to control the type I error from being improperly inflated, the individual two-tailed F value needs to reach $p < .003125$ to be significant. None of the F-tests for the scale ratings reached $p < .003125$ as shown in Table 3. Therefore, there were no differences between the ratings from the web survey and the ratings from the randomly selected large sample using paper-and-pencil survey.

Besides testing the comparability in reliability coefficients and scale means, it was important to test the corresponding inter-scale correlations of different scales in the web sample versus those in the large paper-and-pencil sample just to make sure the correlations between different scales had not been changed by the mode of administration. Significant differences in correlations among different scales in web sample and paper-and-pencil survey would signal that different modes of administration changed the nomological network relationships among different scales (Smith & Leigh, 1997).   Correlation matrices are presented in Table 4. All the correlations are significantly different from zero, which means the scales are related with one another.

The statistical comparisons on whether the correlations in the large paper-and pencil sample differ from the corresponding correlations in the web sample reveal whether the relationships among any pair of scales remain the same through different modes of survey administration. The results of the significance tests on the differences of the correlations in the paper-and-pencil and web surveys are in Table 5. Again, the Bonferroni adjustment was used to control the analysis-wide type I error at .05 when conducting a series of z-tests. Each correlation coefficient (r) was first transformed to Fisher's z_r, then significance tests were conducted on the difference between each pair of z_r between the large paper-and pencil sample and the web sample. Since 28 z-tests were conducted simultaneously to determine whether the corresponding correlation coefficients from these two samples were significantly different, individual two-tailed z value needs to reach $p < .0009$ to control the type I error from being improperly inflated. None of the pair of corrections showed any significant difference between these two samples.

Discussion

Although careful planning went into selecting the large paper-and-pencil sample, the representativeness of the sample was not achieved. Theoretically, an unbiased and representative sample may be achieved by randomly selecting a large sample where every member in the population had an equal chance of being selected into the sample. Our large sample was selected through a one-stage clustering sampling. Classes were randomly selected not individual students. No significant differences were found between the paper-and-pencil sample and the university student population in terms of gender but significant differences were found in race, colleges and number of credit hours enrolled. The comparison between the population and the web sample also showed the same pattern. The web sample was a self-selected convenient sample. Different modes of survey in this study reached different audience and both samples were similar with the population in gender composition but different in race, colleges, and number of credit-hours enrolled. Different modes of survey were known to reach different subsets of the population. This result supports those of Tomsic et al. (2000), in which a detailed measurement of student demographics indicated significant differences in web versus paper-and pencil surveys. Such a result is not necessary a fatal flaw. It provides an opportunity to empirically test the measurement equivalence and survey outcomes when both samples are selected from the same population.

The second research question was designed to empirically test whether there was measurement equivalence between web and paper-and-pencil samples. Comparable reliability coefficients between the two samples indicated that individuals responded to different modes of survey with a similar level of consistency. These results are similar to those of Davis (1999) in that identical measures that were administered via different modes should result in similar levels of internal consistency. All scales in web and paper-and-pencil samples showed almost identical Cronbach's Alpha coefficients for internal consistency. All scales except Parking reached .70 or higher. A subsequent examination of the items and a factor analysis in the Parking scale revealed that the scale included parking availability, cost, and customer service items that were not highly correlated with one another. Thus, Parking scale was excluded from further analysis. The comparison clearly demonstrated that respondents perceived and interpreted the survey items in a similar manner regardless how the items were delivered.

The third research question was designed to test the comparability of quality of service ratings between the two samples. Are results of the web survey similar to the results of the randomly selected large sample? After a significant MANOVA result, a series of F-tests on the service ratings revealed that no significant differences existed between the two samples after Bonferroni adjustment. This indicated that a well-designed survey may generate similar results regardless of the mode of administration. The testing of the differences of correlation coefficients between any pair of scales in web versus paper-and-pencil samples provides an empirical evidence on whether different modes of survey administration change the nomological network relationships among different scales.

After a Bonferroni adjustment was used to control the analysis-wide type I error from being improperly inflated, none of the 28 z values were significant. The results support the inference that modes of survey administration do not change the relationships among different scales.

In summary, the evidence here indicated that a carefully selected large sample might still be different from the population in some demographic variables. Comparisons between a small self-selected web sample and the large sample yielded significant differences in 11 of 14 detailed demographic and academic items. Although the web survey and the paper-and-pencil survey might have reached different audiences, their ratings of the quality of service were similar. The ratings of an attitude survey between the web sample and the large sample from paper-and-pencil survey yielded almost identical reliability coefficients, similar scale means, and similar correlations among different scales.

This finding may be of great interest to university administrators. Acknowledging the time and cost benefits with respect to using the web to administer a survey, the current research suggests that conducting a web survey will not sacrifice the accuracy of the results. Often when utilizing the web for data collection, the measurement equivalence issue is at the heart of the debate. The current research used empirical data to analyze the comparability between the paper-and-pencil and web surveys on student satisfaction of quality of campus services assuming both modes of administrations reaching for the same population during the same time period. This investigation was critical to appropriately develop of a cost efficient way to obtain students' opinions, attitudes and perceptions for universities. The same method may presumably be applied to employees' opinion or attitude surveys in large organizations where internet or intranet access is provided for every employee.

Suggestions for further research would include the use of longitudinal studies and comparisons across different universities. Longitudinal data and multiple-organization data could be used to investigate whether the demographic differences between web and paper-and-pencil samples yield similar patterns or any identifiable trends over time and whether such differences between samples change over time.

References

Booth-Kewley, S., Edwards, J. E, & Rosenfeld, P. (1992). Impression management, social desirability, and computer adminsteration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology, 77*, 562-566.

Buchanan, T. & Smith, J. (1999). Using the Internet for Psychological Research: Personality Testing on the World Wide Web. *British Journal of Psychology*. 90, 125-144.

Corts, D.P., Lounsbury, J.W., Saudargas, R.A., & Tatum, H.E. (2000). Assessing undergraduate satisfaction with an academic department: A method and case study. *College Student Journal, 34,* 399-409.

Cronk, B.C. & West, J.L. (2002). Personality Research on the Internet: A comparison of Web-based and Traditional Instruments in Take-home and In-class settings. *Behavior Research Methods, Instruments & Computers*. 34, 177-180.

Davis, R.N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments & Computers, 31*, 572-577.

Eaton, D. K., Brener, N. D., Kann, L., Denniston, M. M., McManus, T., Kyle, T. M., Roberts, A. M., Flint, K. H., and Ross, J. G. (2010). Comparison of paper-and-pencil versus web administration of the Youth Risk Behavior Survey (YRBS): Risk behavior prevalence estimates. *Evaluation Review, 34*, 137-153.

McMorris, B. J., Petrie, R. S., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2009). Use of web and in-person survey modes to gather data from young adults on sex and drug use: An evaluation of cost, time, and survey error based on a randomized mixed-mode design. *Evaluation Review, 33*, 138-158.

Schmidt, W.C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, 29, 274-279.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* New York, NY: Houghton Mifflin

Smith, M., & Leigh, B. (1997). Virtual subjects: Using the internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers*, *29*, 496-505.

Stanton, J. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology, 51*, 709-725.

Tomsic, M.L, Hendel, D.D., & Matross, R.P. (2000). *A world wide web response to student satisfaction surveys: Comparisons using paper and internet formats.* Paper presented at the 40th annual meeting of the Association of Institutional Research, Cincinnati, OH.

Vispoel, W.P. (2000). Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. *Measurement & Evaluation in Counseling & Development, 33*, 133-144.

Table 1. *Population and Sample Characteristics*

Gender	CSU student population		Paper-and Pencil sample		Web Sample	
	Observed Frequency	%	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
Male	7020	43.88	887	845.14	96	104.00
Female	8978	56.12	1039	1080.86	141	133.00
			$\chi^2 = 3.695$	$p = .055$	$\chi^2 = 1.097$	$p = .295$
College						
A & S	5540	35.31	760	724.90	108	80.15
Business	3265	20.81	460	427.22	50	47.24
Education	2842	18.11	355	371.87	28	41.12
Engineering	1036	6.60	239	135.56	22	14.99
Law	816	5.20	102	106.77	1	11.81
Urban	669	4.26	86	87.54	14	9.68
University Studies	1522	9.70	51	199.15	4	22.02
			$\chi^2 = 194.36$	$p = .000$	$\chi^2 = 43.861$	$p = .000$
Race						
White	9838	67.86	1439	1393.12	173	150.64
Black	2897	19.98	341	410.23	18	44.36
Hispanic	418	2.88	72	59.19	4	6.40
Asian	437	3.01	121	61.88	12	6.69
Native American	36	0.25	13	5.10	0	0.55
Foreign	872	6.01	67	123.48	15	13.35
			$\chi^2 = 110.52$	$p = .000$	$\chi^2 = 24.851$	$p = .000$
# credit hours enrolled						
4 and under	2887	18.05	134	393.04	21	42.59
5 to 8	3297	20.61	287	448.86	37	48.64
9 to 12	3484	21.78	507	474.32	78	51.40
13 to 16	5019	31.37	895	683.30	74	74.04
17 and over	1311	8.19	355	178.48	26	19.34
			$\chi^2 = 471.51$	$p = .000$	$\chi^2 = 29.789$	$p = .000$

Table 2. *Comparability of Cronbach α Coefficients for Web and Paper-and-Pencil Scales*

	Items	Web Survey α (Valid N)	Paper-and-Pencil Survey α (Valid N)
Bursar	14	.88 (79)	.88 (513)
Financial Aid	11	.87 (127)	.87 (950)
Parking	9	.54 (155)	.49 (1190)
Registrar's	12	.83 (78)	.88 (636)
Food Services	8	.87 (202)	.89 (1600)
Advising	8	.92 (123)	.93 (1177)
Bookstore	7	.64 (161)	.74 (1355)
Admissions	8	.92 (104)	.88 (1076)
Overall University Experience	4	.86 (205)	.86 (1625)

Note: The numbers in parentheses represent the valid sample size for each scale. The valid N differed due to list-wise deletion of missing values. Some items were deleted to increase the reliability coefficients.

Table 3. *Comparisons of Scale Means in the Quality of Service Survey between Web and Paper-and-Pencil Administration*

	Web	Paper-and-Pencil	F-test	p-value
Bursar	3.35 (.79)	3.43 (.79)	1.146	.285
Financial Aid	3.20 (.88)	3.32 (.85)	2.681	.102
Registrar's	3.50 (.71)	3.49 (.78)	0.032	.859
Food Services	2.94 (1.00)	3.16 (.97)	6.315	.012
Advising	3.52 (1.11)	3.57 (.96)	0.289	.591
Bookstore	3.17 (.60)	3.27 (.68)	1,268	.093
Admissions	3.37 (.95)	3.54 (.85)	5.090	.024
Overall University Experience	3.56 (.91)	3.42 (.92)	3.114	.078

Note: The number in parentheses represents the standard deviation for that scale. Bonferroni adjustment requires each individual t-test to reach $p < .003125$ to be significant.

Table 4. *Correlation Matrices among Scales in the Paper-and-Pencil Sample versus Web Sample*

Scale	1	2	3	4	5	6	7	8
			Paper-and-Pencil Sample					
1. Bursar	1.00							
2. Financial Aid	.582**	1.00						
3. Registrar's	.579**	.508**	1.00					
4. Food Services	.393**	.364**	.386**	1.00				
5. Advising	.418**	.400**	.481**	.315**	1.00			
6. Bookstore	.414**	.391**	.431**	.440**	.430**	1.00		
7. Admission	.526**	.449**	.567**	.413**	.524**	.438**	1.00	
8. Overall	.453**	.380**	.455**	.382**	.446**	.364**	.482**	1.00
			Web Sample					
1. Bursar	1.00							
2. Financial Aid	.515**	1.00						
3. Registrar's	.570**	.512**	1.00					
4. Food Services	.272**	.340**	.313**	1.00				
5. Advising	.302**	.270**	.336**	.277**	1.00			
6. Bookstore	.393**	.351**	.466**	.282**	.248**	1.00		
7. Admission	.331**	.366**	.477**	.329**	.316**	.257**	1.00	
8. Overall	.328**	.292**	.401**	.299**	.453**	.260**	.453**	

Table 5. The *p-Values of the Differences of Corresponding Correlation Coefficients in the Paper-and-Pencil Sample versus Web Sample*

Scale	1	2	3	4	5	6	7
1. Bursar							
2. Financial Aid	.214						
3. Registrar's	.846	.944					
4. Food Services	.055	.729	.241				
5. Advising	.062	.065	.014	.571			
6. Bookstore	.716	.548	.524	.010	.004		
7. Admission	.002	.240	.107	.220	.001	.007	
8. Overall	.032	.198	.337	.185	.902	.093	.628

Bonferroni adjustment requires each individual z-test to reach $p < .0009$ to be significant.