

From Paper to Pixels: The Effects of Modality on End-of-Course Survey Responses

Alexander Cohen, Ph.D.

Ashford University San Diego, California, United States
Email: alexander.cohen@ashford.edu

Jennifer V. Irwin, Ph.D.

Ashford University San Diego, California, United States

Edward Gall, M.S., M.Div.

Ashford University San Diego, California, United States

Abstract

This study examines how end-of-course surveys administered at the same institution in two modalities (online and in-class) differ. In particular, substantive responses to qualitative questions appeared systematically less substantive than those administered in class. However, students did respond to qualitative questions at a higher rate, even though these responses were, on average, less substantive. It also appears that survey response was far less consistent with one another in the online modality than the traditional, in-class modality. On the other hand, there is very little convincing evidence that shifting to online end-of-course surveys leads to more positive or negative responses.

Keywords *end-of-course surveys, assessment, online learning, modality shifts, institutional assessment*

INTRODUCTION

End-of-course surveys are important. Throughout academia, they are routinely utilized to assess learning, guide institutional development, address instructor efficacy, and understand how courses contribute to student learning. Consequently, it is important to understand how the method of survey delivery can affect survey response. This study focuses on the effects of moving from a traditional, in-class, pen-and-paper survey to the online survey administration. As institutions continue to examine ways to cut costs, reduce the ecological impact of paper usage, and more accurately evaluating their faculty and programs, obtaining a better understanding of how modality can affect survey response is an important contribution. This article represents one such endeavor.

This study examines the difference between student responses to end-of-course surveys administered at the same institution in sequential years. In 2013, end-of-course surveys were administered in a traditional, in-class format; in 2014, they were administered entirely online and were near-compulsory. This provides an important lens into the effects of varying the modality of survey administration in a University-wide, large-N setting.

This article is organized in the following sections. Section One provides a brief review of the literature on the relationship between survey modality and survey response. In Section Two, we describe a natural experiment at our institution, wherein the modality of end-of-course surveys was unilaterally switched

between Academic Calendar years. We discuss, in detail, the particularities of both methods of survey administration and the data that we will analyze. In Section Three, we generate hypotheses regarding the relationship between survey modality and survey response. In Section Four, we provide analysis. Section Five concludes with a discussion.

THE RELATIONSHIP BETWEEN SURVEY MODALITY AND SURVEY RESPONSE

Evaluating the student learning experience is an important cornerstone of collegial academics. After all, if we do not know where our students are succeeding and struggling, then we cannot continue to refine both craft and curriculum. End-of-course surveys are one important—and reasonably systematic—measure of student learning, and are regularly utilized to assess the student experience, course quality, and instructor quality (e.g., Richardson, 2005). As universities are increasingly required to cut budgets, administering end-of-course surveys online has emerged as an increasingly attractive option due to lower costs associated with administration, prompting an increase in their use (Anderson, Brown, & Spaeth, 2006). Data can be collected without scanning paper forms, eliminating work hours necessary for scanning and, in some cases, typing out qualitative comments to ensure their anonymity. The consequences of this shift have been the subject of interesting scholarly discussion.

There are many reasons to anticipate systematic differences between online and in-class surveys. For instance, the attractiveness of the instructor can positively influence ratings of instruction (Rocca & McCroskey, 1999), suggesting that there are strong effects related to administering a survey in a traditional classroom. Likewise, Web-based surveys offer students an increased sense of anonymity and may be associated with greater and more honest self-disclosure (Conboy, Domar, & O'Connell, 2001). Wyatt (2000) examines how end-of-course surveys are conceptualized by students, argues that in-class and online surveys are perceived differently by participants, which leads to different sorts of responses. Lazar and Preece (1999), on the other hand, believe that in-class and online surveys provide relatively parallel results, so long as the format is similar.

Empirical tests of differences between survey responses across modalities have yielded mixed results. Some scholars have failed to find significant empirical differences between online and in-class surveys (Dommeyer et al 2002, 2004; Layne, DeCristoforo, & McGinty, 1999). Others have pointed out that such null results may be the result of methodological issues, arguing that simpler bivariate comparisons ignore other possible causal factors (Krautmann & Sander, 1999). Key factors include time of day, class size, class level (Millea & Grimes 2002), individual GPA and student major (Nowell, 2007), and instructor-specific factors (Marsh & Roche, 1999). Drawing from data across 28 separate courses and 15 faculty members, Nowell, Gale, and Handley (2010) do control for a variety of factors and demonstrate that online end-of-course surveys¹ administered online tend to exhibit lower ratings than those administered in class.

Other scholars have offered complementary contributions to the narrative. While Venette, Sellnow, and McIntyre (2010) do not find any differences in responses in online and in-class surveys, they note a greater level of descriptive detail in online surveys. Others have focused on the lower response rates associated with transitioning to online surveys, and noted how the exclusion of previously included subpopulations can introduce bias into results (Sax, Gilmartin, & Bryant, 2003).

In short, the effects of moving from in-class to online surveys are still being understood. Such effects are complex, and appear contingent upon a variety of factors, and so it behooves scholars of higher education to examine the question from different perspectives. Consequently, studies may obtain differing results because they account for different factors and make use of different samples. Studies are also focused on different dependent variables. In this particular study, we leverage a large-N comparison of survey response across different modalities in sequential years, and look at how the shift in modality affected at two dependent

¹ While we use the term end-of-course survey or EOCS here for the sake of consistency, other terms are often applied to the same thing, such as “student evaluations of teaching” (SET).

variables that are prominent in the literature: the positivity of quantitative and qualitative responses, and the depth of qualitative responses. We also test a somewhat novel idea: that online surveys may be associated with more ‘random’ responses. In the next Section, we provide a detailed description of the data that we will analyze.

DATA

Data was collected from Ashford University’s campus in Clinton, Iowa at the close of two different semesters: Spring of the 2013 Academic Year and Spring of the 2014 Academic Year. At its core, the campus was a traditional liberal arts environment focused on offering small courses and a variety of degrees across its four Colleges. In 2013, survey administration was conducted through traditional pen-and-paper surveys. Surveys were distributed to instructors with instructions to distribute in-class any time within the final two weeks of the semester. While instructors were free to differ in their approach and in-class instructions regarding completing the surveys, they were required to leave the room while surveys were completed by students. Many instructors chose to hand the surveys out at the end of the classroom period, though some distributed them at the start of class, left the room for a predesignated time, then returned to hold class. When complete, students would give surveys to a volunteer student, who placed them in an opaque folder and brought them directly to the Administrative Assistant to their College. At that point, they were sent to Assessment staff and processed. Surveys consisted of a battery of questions asking students to rate various aspects of course and instructor on a Likert scale, as well as an open-ended qualitative question. Quantitative data was scanned into a single digital dataset through SCANTRON technology, and qualitative responses were manually typed into fields of this dataset by Assessment personnel. Please see the Appendix for question wording.

In Spring of 2014, the University shifted to online end-of-course surveys in order to eliminate the cost associated with manually scanning quantitative and qualitative student responses, and to reduce the environmental impact of utilizing so much paper. When students attempted to access the online portal any of their courses during the last week of class, the survey immediately filled their screen, and they were required to complete it in order to access their course’s online resources, which included gradebook data, assignments, and study guides. The online portal is a tool hosted by Pearson’s eCollege Learning Management System, which was used by nearly all University faculty² to organize student grades, host asynchronous discussion, facilitate the submission of assignments and final papers, and provide links to the library and other necessary resources. Nearly all students did check their online portals during this window, as many courses required online submission of final materials, and students were eager to see their grades calculated prior to the final. Response rates, as a result, climbed to 98%. We thus regard this sort of online survey administration as ‘near-compulsory’, as every student who accessed their online portal for a given course was forcibly exposed to an end-of-course survey and was required to complete that survey to access desired materials. Once the survey appeared on their screen, students were required to at least fill in the radio buttons for all questions in the quantitative battery before clicking on a ‘submit’ button, after which the survey vanished and the usual course materials appeared.

Two items are worth noting. First, the Spring semesters of 2013 and 2014 were selected for two reasons: to hold possible seasonal effects constant, and because a similar battery of courses were offered in both semesters, which reduces variation in question response. Second, unfortunately, the question wording of both qualitative and quantitative questions did shift between 2013 and 2014. While the quantitative questions all make use of the same 5-point Likert scales, such that higher numbers indicate to more positivity regarding course and instructor, question wording does shift across surveys. There was one qualitative question in 2014

² Several faculty opted out of using the online system, particularly in independent studies and music practica. Students in courses or sections that did not utilize an online course site account for less than .5% of the total courses and sections offered in this semester.

and two in 2015, and the wording in both is different. Such shifts, obviously, are not ideal in a perfectly controlled study. However, as we discuss during our analysis, the changes in question wording should not have a powerful effect on many of our dependent variables. Additionally, the quantitative responses to survey responses in each year are computed as a mean of responses to each of the thirteen questions asked; the aggregated nature of the data that we analyze should strongly depress any framing effects caused by individual questions. Therefore, we feel that the large-N nature of this study makes up for some of these shortcomings, and that by noting the issue and providing the questions themselves, this study nevertheless makes a contribution to a rich and vibrant discussion.

THEORY

This study focuses on the effect of shifting modalities of end-of-course surveys on five dependent variables: the positivity of quantitative responses, the positivity of qualitative responses, the substantive depth of qualitative responses, the quantity of qualitative responses, and the randomness of quantitative responses.

There are several general reasons why the modality of end-of-course surveys should impact responses. First and foremost, the environment of survey administration is drastically different across modalities. In traditional survey administration, students are gathered in the same room and must record their thoughts with a pen or pencil. Nothing prevents them from turning in blank surveys. When surveys are administered at the start of class, students are given a designated time with little to do but think of the survey; when surveys are administered at the end of class, students may speed through responses to exit the room quickly, though their decision may be impacted by the relative diligence—or lack thereof—of their peers. The environment is not particularly rushed; while finals may loom, students are in class at a time scheduled for that purpose. They are also more immediately impacted by the instructor, who is at least present when the surveys are handed out, and who may even give a brief overview about how student input is valuable for improving the quality of the course (and how it may be used in their own evaluations by their Dean). A strong social factor is also at work: students may talk while they fill out surveys to complain about the class or compliment it. Less conscientious students may watch as more conscientious students diligently think about the survey and take significant time to complete it, which may pressure them into doing the same.

Taking an end-of-course survey online is a very different experience. There is no social component, as students completing the survey must do so in isolation from other students. The instructor is not directly present. The time of day differs. The physical environment of administration is wildly variable. Some students may take it in a computer lab, others in their dorms late at night, and others while visiting their parents. Additionally, the particular method of online survey administration in this population introduces an important factor. Whenever students attempted to log into their online course portals during the last week of class, they were forced to fill in the radio buttons on the survey. It is conceivable that many students logged into their courses to hurriedly turn in a final paper, collect course resources, or nervously check their grades, and thus had little desire to take an online survey about the course. Further, the near compulsory nature incorporated a new population of students responded to surveys: those who simply turned in blank surveys in-class, or walked out of the room without completing them. All of this indicates that, generally speaking, there ought to be differences in survey response. Whether such differences are measurable and systematic is the subject of this study.

More specifically, we hypothesize that surveys administered in an online, near-compulsory format should be systematically less positive than those administered in class. We know from the differential in response rates between the two modalities of survey administration that the increased number of students responding to near-compulsory online surveys included students who did not turn in surveys administered in class, either because they were in a hurry to leave class, did not want to bother providing feedback, or were absent from class on the day of survey administration. Throughout analysis, we call this population of students “Survey Avoiders”. We postulate, but cannot confirm, that based on such behaviors Survey Avoiders

may be underperformers or otherwise unhappy with the course, but do not reflect such dissent until mandated by the online portal. It is also possible that students, anxiously accessing their student portals seeking to access materials of use, were frustrated by the appearance of a mandatory screen and took their frustration out on the survey questions. Finally, the lack of recent interaction with the instructor (some of whom even bring donuts on evaluation day) may have a negative impact on online response.

We also hypothesize that the responses to qualitative questions should exhibit greater depth when end-of-course surveys are administered in class. First, the addition of Survey Avoiders to the population of respondents in the online population could result in qualitative responses that are pithy rather than substantive. After all, these are students who either do not fill out in-class end-of-course surveys, or miss class when they are administered. It is reasonable to assume that their inclusion in the sample might increase the number of responses such as “good class” or “you suck.” Second, in-class end-of-course surveys should ideally take place in a reflective environment, in the immediate context of class, with some instructor preparation and social pressures that might induce students to take their time and provide good, substantive qualitative responses to survey questions. Finally, even well-meaning students rushing to access their course portals may ignore prompts for qualitative feedback because they are anxious to access course materials. Our coding procedures regarding the ‘substantivity’ of qualitative comments is discussed in sections below.

Of course, even pithy qualitative comments provide some information, and so it is also important to examine the rate of qualitative response across modality. We hypothesize that the rate of qualitative responses will decrease in online surveys for the above reasons.

Finally, we hypothesize that students who took the online survey will display a more random pattern of response. Our logic here is similar that described above: Survey Avoiders and those rushing to access course materials may simply make ‘Swiss cheese’ out of the evaluations and fill them out randomly.

In summary, our testable hypotheses are:

H₁: Students taking a near-compulsory online end-of-course survey were more likely to respond to questions in negative ways than students whose responses were collected through in-class surveys, in both qualitative and quantitative responses.

H₂: Students taking a near-compulsory online end-of-course survey were more likely to provide non-substantive responses to qualitative questions than students whose responses were collected through in-class surveys.

H₃: Students taking a near-compulsory online end-of-course survey were less likely to provide any qualitative responses than students whose responses were collected through in-class surveys.

H₄: Students taking a near-compulsory online end-of-course survey were more likely to provide a more random pattern of quantitative responses than students whose responses were collected through in-class surveys.

ANALYSIS

In the section below, we examine each hypothesis in turn.

Testing Hypothesis One. Recall, H₁: Students taking a near-compulsory online end-of-course survey were more likely to respond to questions in negative ways than students whose responses were collected through in-class surveys, in both qualitative and quantitative responses.

We examine, first, the quantitative responses. In each survey, students responded to a battery of 13 quantitative questions about the course and instructor. All responses are measured on a five-point Likert scale where 1 = Disagree Strongly, 2 = Disagree, 3 = No Opinion, 4 = Agree and 5 = Agree Strongly. While question wording was different in each year, positive responses to questions always correspond to positive evaluations of some element of the course or its instructor. Again, we average responses to each question into a single mean score for each student in each modality to limit variation caused by variance in question wording across modalities.

The mean of all responses in the online modality is 4.221 (SD = .984; N = 2950), and the mean for in-class surveys is nearly identical, at 4.3 (SD = 1.32; N = 3362)³. An unpaired t-test conducted on these samples shows that the difference is statistically significant ($t(38453) = -11.21, p = .000$, two-tailed). However, its substantive impact appears negligible.

Next, we compare the average scores of each instructor who taught courses before and after the transition to online surveys. By comparing mean responses at the instructor-level, we are able to control for variation caused by differing levels of instructor quality. Because not all instructors taught in both years due to staffing changes, not all instructors from the previous analysis are included. The table below displays, at the instructor-level, the mean scores across all online and in-class questions, the difference between these means, and the statistical significance of those differences. As a whole, responses to online surveys were .136 points higher, on average, for each instructor than those collected from in-class surveys. However, as Table 1 shows, a t-test of each relationship at the instructor level shows that few differences between mean scores achieve statistical significance, which can be, in part, attributed to instructors who taught very few students, thereby yielding small subsample sizes. When only statistically significant differences between online and in-class scores are examined, the positive effect appears stronger: for these instructors, mean scores went up in the online modality by .321.

Table 1 Results of Difference of Means Tests Conducted on EOCS Results for Instructors Teaching both before and after Shift to Online Survey Administration, 2013-2014.

Instructor	In-class mean	Online mean	Dif. of means	Instructor	In-class mean	Online mean	Dif. of means
1	3.98	4.77	0.79***	45	4.08	4.302	0.222
2	4.14	4.5	0.36***	46	3.585	4.134	0.549
4	4.21	4.39	0.18	47	3.78	3.75	-0.03
6	4.28	4.36	0.08	48	4.143	3.469	-0.674***
7	4.45	4.32	-0.13	50	4.315	4.125	-0.19
12	3.81	3.57	-0.24	52	3.523	3.977	0.454**
13	3.71	3.77	0.06	53	4.44	4.42	-0.02
14	4.05	3.55	-0.5**	54	4.22	4.5	0.28
15	4.3	4.6	0.3***	55	4.48	4.42	-0.06
16	3.93	4.09	0.16	56	4.284	3.965	-0.319
17	4.37	4.41	0.04	58	4.27	4.521	0.251*
18	4.19	4.35	0.16	59	3.65	4.35	0.7*
22	4.11	4.198	0.088	60	4.2	5.47	1.27***
23	4.14	4.498	0.358*	63	4.06	4.558	0.498***
27	3.81	4.24	0.43***	64	4.034	3.904	-0.13
28	3.73	3.43	-0.3	66	4.174	3.958	-0.216
29	3.96	3.74	-0.22	67	4.279	4.57	0.291
30	3.65	3.47	-0.18	68	4.201	3.934	-0.267*
32	4.278	4.438	0.16	70	3.953	4.306	0.353
34	4.22	4.455	0.235	72	4.164	4.075	-0.089

³ Note that while the absolute number of responses was actually larger among in-class surveys, the response rate was dramatically higher for online surveys. This was because a large-scale reduction in the student body and the offering of fewer classes in the 2014 Academic Year, when online surveys were instituted.

35	4.04	4.641	0.601***	73	4.407	4.684	0.277*
36	3.85	4.122	0.272	74	4.37	4.48	0.11
37	4.12	4.544	0.424***	75	3.48	3.355	-0.125
39	3.916	3.869	-0.047	76	3.939	4.357	0.418***
41	4.26	3.81	-0.45***	78	4.136	4.769	0.633***
43	4.088	4.359	0.271	79	4.37	4.604	0.234***
				80	4.42	4.036	-0.384

* = p<.05, ** = p<.01, *** = p<.001. Significance calculated as the two-tailed p-value computed using the t-distribution; in other words, the probability of observing a greater absolute value of t under the null hypothesis.

A third way to test this relationship is through Ordinary Least Squares regression analysis. Here, we regress the average quantitative responses of on each end-of-course survey⁴ on a dummy variable, *pixel*, coded 1 if the survey was administered online. Here, we also control for instructors through ninety dummy variables, coded 1 to indicate the instructor being evaluated. This helps, at least in part, to separate trends that may be due to instructor expertise rather than the modality of survey. This regression, shown below, demonstrates that, indeed, students who responded online tended to provide more positive reviews, though the substantive impact is minimal at an increase of just .13.

Table 2 Results of OLS Regression of Online Survey Administration on EOCS Responses, 2013-2014.

Variable	Coefficient	Robust SE	P> t
Pixel	.13	.021	.000
... 90 dummy variables ...	---	---	.000
Constant	3.947	.24	.000

N = 5494

All three tests suggest that, at least in this context, transitioning to near-compulsory online end-of-course survey administration was associated with a statistically significant but substantively small positive impact on survey response.

We turn, next, to a comparison of the relative positivity of qualitative responses in the online and in-class surveys. Unfortunately, here, the changes in question wording across surveys are far more impactful on results. With so few questions, averaging them to create an aggregate figure does not help ‘wash out’ variance caused by question wording. And, in the online version of the survey, the two questions asked were quite pointed, with one likely to produce positive responses, and another likely to produce negative responses.

We coded qualitative responses as ‘1’ if they were positive, and ‘0’ if negative. Responses that were a mix of both positive and negative, given no value and omitted from analysis. Blank responses were omitted from this analysis. See Table 3 below for results. Responses from the Spring 2013 survey were coded by one researcher, while responses from the Spring 2014 survey were coded by another (both using the same coding method). Early in the coding process, the two researchers met to ensure consistency and reliability. Only one open-ended question was asked in Spring 2013, while two were asked in Spring 2014. The Spring 2013 question was “Your comments regarding your instructor, course, and Ashford University are welcome”. On the Spring 2014 survey, students were given two open-ended questions: (1) “What were the strengths of this course in preparing you for future courses and/or your career?”, and (2) “How could the course be changed or improved to better prepare you for your career?”.

⁴ Note that because Likert scores are averaged across all 14 categories, this is a continuous variable that does not violate OLS’s basic assumptions.

Table 3. Mean ‘Positivity’ of Qualitative Responses, 2013 and 2014.

Year	Question	N	Mean ‘Positivity’ Score	Standard Error (Mean)
2013	Your comments?	1288	.68	.013
2014	Strengths of course?	1111	.92	.008
2014	How could course be improved?	923	.34	.016

It is difficult to take much from these results, largely due to the powerful framing effect of the questions asked. It is perhaps worth noting how many more students were willing to answer the question regarding course improvement, as opposed to questions about the course’s strengths, in 2014.

In the balance, then, this section suggests that moving to an online form of survey administration had, at best a modest effect on survey response. When we take into account the possible influence of question wording effects on this outcome, it is difficult to make much of these trends.

Testing Hypothesis Two and Three. Recall H₂: Students taking a near-compulsory online end-of-course survey were more likely to provide non-substantive responses to qualitative questions than students whose responses were collected through in-class surveys.

In both populations, we first coded responses as ‘substantive’ if students provided meaningful responses (e.g., “The course was average at best. Positive-service was fun and helpful, professor was great to talk to, class was professional. Negative-textbook needs changing, service should be worth more, ...”); as opposed to non-substantive (e.g., “great professor!”). Responses that were substantive in nature were coded as ‘1’ and non-substantive responses were coded as ‘0’. Blank responses were omitted from this analysis. Here, the variation in questions across surveys is of minimal concern, as there is little reason to suspect that the wording of the qualitative questions would lead to differing levels of substantivity in responses. As Table 5 shows, the substantivity of responses given to the online end-of-course survey were dramatically lower than those administered in class.

Table 5 Mean ‘Positivity’ of Qualitative Responses, 2013 and 2014.

Year	Question	N	% of Total Students Providing Qualitative Responses	Mean ‘Substantivity’	Standard Error of Substantivity (Mean)	Two-Tailed Significance Of Mean
2013	Your comments?	1174	34.9%	.76	.011	.000
2014	Strengths of course?	1219	41%	.24	.004	.000
2014	How could course be improved?	1112	37%	.34	.008	.000

In order to further verify these results, we conducted a t-test for the difference of means between the 2013 results, and both 2014 results. The p-value is .002, suggesting that, indeed, responses were far less substantive when provided by respondents to the online survey. We did not conduct analysis at the instructor level because qualitative responses are relatively infrequent, leading to very small sample sizes in one or both years.

Next, recall H₃: Students taking a near-compulsory online end-of-course survey were less likely to provide any qualitative responses than students whose responses were collected through in-class surveys. The

table above demonstrates that, contrary to expectations, students were more likely to provide *any* responses to qualitative questions posed in near-compulsory end-of-course surveys. However, in the balance, such responses are far less substantive.

Testing Hypothesis Four. Finally, recall H₄: Students taking a near-compulsory online end-of-course survey were more likely to provide a more random pattern of quantitative responses than students whose responses were collected through in-class surveys.

Thus, we test if greater randomness exists in responses to high response rate online surveys, relative to in-class surveys. The simplest snapshot here is to first average the standard deviation in aggregate responses to each survey question as a measure how consistent question responses are to one another, and then compare the average of these averages across survey types. The average standard deviation across all online surveys is .988, while the average standard deviation across all in-class surveys is .905. Answers to questions posed in the online survey indeed display a higher standard deviation than questions administered on papers.

This can further be broken down at the instructor level. In Table 5 below, we see that the standard deviation in quantitative responses to end-of-course survey questions is greater in all but three cases, and many of these reach levels of statistical significance.

Table 6 Differences in Standard Deviation across Quantitative EOCS Responses, Instructor-Level, 2013 and 2014

Instructor	In-class SD	Online SD	Dif. Between SDs	Instructor	In-class SD	Online SD	Dif. of means
1	0.452	0.619	-0.167	45	0.167	0.081	0.086
2	0.625	0.518	0.107	46	0.921	0.716	0.205
4	0.733	0.414	0.319	47	0.996	0.554	0.442
6	0.603	0.399	0.204	48	1.086	0.562	0.524
7	0.651	0.41	0.241	50	0.714	0.457	0.257
12	0.197	0.083	0.114	52	0.945	0.517	0.428
13	0.332	0.063	0.269	53	0.629	0.362	0.267
14	0.865	0.531	0.334	54	0.869	0.439	0.43
15	0.694	0.358	0.336	55	0.596	0.212	0.384
16	0.745	0.601	0.144	56	1.33	0.316	1.014
17	0.0863	0.0431	0.0432	58	0.589	0.468	0.121
18	0.112	0.053	0.059	59	0.951	0.514	0.437
22	0.875	0.544	0.331	60	0.604	0.594	0.01
23	0.732	0.482	0.25	63	0.421	0.422	-0.001
27	0.775	0.652	0.123	64	0.823	0.431	0.392
28	1.092	0.654	0.438	66	0.937	0.454	0.483
29	1.28	0.485	0.795	67	0.56	0.514	0.046
30	0.748	0.526	0.222	68	0.708	0.412	0.296
32	0.588	0.478	0.11	70	0.846	0.542	0.304
34	0.752	0.444	0.308	72	0.891	0.426	0.465
35	0.455	0.494	-0.039	73	0.428	0.317	0.111
36	0.795	0.524	0.271	74	0.556	0.366	0.19
37	0.642	0.582	0.06	75	1.09	0.684	0.406
39	0.984	0.571	0.413	76	0.733	0.612	0.121

41	0.899	0.427	0.472	78	0.502	0.392	0.11
43	0.639	0.472	0.167	79	0.677	0.338	0.339
				80	1.203	0.466	0.737

In conclusion, at both the instructor level and in the aggregate, the standard deviation of responses clearly increased as a result of moving to near-compulsory online surveys. This suggests that students were more likely to randomly fill out bubbles when forced to input an end-of-course survey online, than when voluntarily asked to do so in class.

DISCUSSION

This analysis suggests that, at least in this environment, there were systematic differences between administering end-of-course surveys in the online modality, relative to traditional pen-and-paper administration. In particular, substantive responses to qualitative questions appeared systematically less substantive than those administered in class. However, students did respond to qualitative questions at a higher rate, even though these responses were, on average, less substantive. It also appears that survey response was far more random in the online modality than the traditional, in-class modality. On the other hand, there is very little convincing evidence that shifting to online end-of-course surveys leads to more positive or negative responses.

These results are important to note as institutions of various size and complexity will continue to shift towards online end-of-course surveys as a mechanism to reduce cost. Such surveys will remain important tools through which to assess pedagogy, and so it important to understand their strengths and limitations. Moreover, this study of this particular institution’s journey in this respect may provide guidance to similar institutions.

Of course, this is just one case of many, and is not without its limitations. Shifts in question wording across sample sizes, though mitigated through aggregating quantitative scores, may have affected some statistical inferences. Additionally, where variations in survey response do exist, we cannot be certain of the mechanism. Whether the near-compulsory nature of these end-of-course surveys or their digital modality (or some combination of the two) *caused* differences in outcomes is impossible to say with this data. More exploration here would be instructive. Nevertheless, despite its limitations, this study makes an honest contribution to the growing body of literature on this topic.

REFERENCES

- Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. *Innovate*, 2(6). Retrieved from <http://innovateonline.info/index.php?view=article&id=301>
- Conboy, L., Domar, A., & O’Connell, E. (2001). Women at mid-life: Symptoms, attitudes, and choices, an internet-based survey. *Maturitas*, 28, 129-136.
- Dommeyer, C.J., Baum, P., Chapman, K., & Hanna, R.W. (2002). Attitudes of business faculty toward two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education*, 27(5), 445-462.
- Dommeyer, C.J., Baum, P., Hanna, R.W., & Chapman, K.S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education*, 29(5), 611-623.
- Krautmann, A.C., & Sander, A. (1999). Grades of student evaluation of teachers. *Economics of Education Review*, 18(3), 59-63.
- Layne, B.H., DeCristoforo, J.R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2), 221-232.

- Lazar, J., & Preece, J. (1999). Designing and implementing web-based surveys. *Journal of Computer Information Systems*, 39(4), 63-67.
- Marsh, H.W., & Roche, L.A. (1999). Rely on SET research. *American Psychologist*, 54(7), 517-518.
- Millea, M., & Grimes, P.W. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, 36(4), 582-591.
- Nowell, C. (2007). The impact of relative grade expectations on student evaluation of teaching. *International Review of Economics Education*, 6(2), 42-56.
- Nowell, C., Gale, L., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment and Evaluation in Higher Education*, 34(4), 463-475
- Richardson, J. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment and Evaluation in Higher Education*, 30(4), 387-415.
- Rocca, K.A., and J.C. McCroskey. (1999). The interrelationship of student ratings of instructors' immediacy, verbal aggressiveness, homophily, and interpersonal attraction. *Communication Education*, 48, 308-316.
- Sax, L.J., Gilmartin, S.K., & Bryant, A.N. (2003). Assessing response rate and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44(4), 409-432.
- Vennete, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment and Evaluation in Higher Education*, 35(1), 101-115.
- Wyatt, J.C. (2000). When to use web-based surveys. *Journal of the American Medical Informatics Association*, 7, 426-430.

APPENDIX

Ashford University – Campus End-of-Course Student Survey Questions (In-Class, Paper Version – Spring 2013)

1. The faculty member was professional in the presentation of the course (punctual, appropriate attire, maintenance of discipline).
2. The faculty member showed organization (was prepared for class, followed the syllabus, and did not routinely venture off on tangents).
3. The faculty member was adaptable (flexible, covered course material using a variety of techniques, was available outside of class).
4. The grading in this course was fair and consistent (definite grading criteria, objective basis for grades, and no favoritism).
5. The faculty member stimulated interest in the subject (showed enthusiasm for and knowledge of the subject matter).
6. The textbook was appropriate for the level at which this course was offered.
7. Other materials complemented the course (films, handouts, and outside readings).
8. Activities in this course seemed appropriate for the type of course content (presentations, field, trips, projects, etc.).
9. This course has provided preparation to continue with additional courses in this topic.
10. This course has provided preparation for work in a related career.
11. This course would prove useful to other disciplines.
12. The material in this course was challenging.
13. This course was too difficult for my capabilities.

*OPEN-ENDED RESPONSE: Your comments regarding your instructor, course, and Ashford University are welcome.

Ashford University – Campus End-of-Course Student Survey Questions (Online Version – Spring 2014)

1. The instructor promotes active classroom participation of students.
2. The instructor fosters critical thinking throughout the course.
3. The instructor adds their perspective, such as knowledge and experience, to the course content.
4. The instructor communicates and promotes high expectations.
5. The instructor's feedback aligns with their communicated expectations.
6. The instructor provides feedback in a timely manner.
7. The instructor provides useful feedback for improving student's quality of work.
8. The instructor provides consistent grading across assignments.
9. I would recommend this instructor to another student.
10. Course assignments require me to think critically.
11. The course content (assignments/readings/study materials) is engaging.
12. The course content and activities will help me in my career.
13. I would recommend this course to another student.

Optional Open Ended Questions:

14. What were the strengths of this course in preparing you for future courses and/or your career?
15. How could the course be changed or improved to better prepare you for your career?