The effect of incentives and other instructor-driven strategies to increase online student evaluation response rates

James Goodman, Robert Anson & Marcia Belcheir

To cite this article: James Goodman, Robert Anson & Marcia Belcheir (2015) The effect of incentives and other instructor-driven strategies to increase online student evaluation response rates, Assessment & Evaluation in Higher Education, 40:7, 958-970, DOI: 10.1080/02602938.2014.960364

To link to this article: https://doi.org/10.1080/02602938.2014.960364

Published online: 22 Sep 2014.

Submit your article to this journal

Article views: 858

View Crossmark data

Citing articles: 11 View citing articles
The effect of incentives and other instructor-driven strategies to increase online student evaluation response rates

James Goodman a*, Robert Anson b and Marcia Belcheir c

aDepartment of Music, Boise State University, Boise, ID, USA; bInformation Technology and Supply Chain Management, Boise State University, Boise, ID, USA; cOffice of Institutional Research, Boise State University, Boise, ID, USA

Course evaluations (often termed student evaluations of teaching or SETs) are pervasive in higher education. As SETs increasingly shift from pencil-and-paper to online, concerns grow over the lower response rates that typically accompany online SETs. This study of online SET response rates examined data from 678 faculty respondents and student response rates from an entire semester. The analysis focused on those tactics that faculty employ to raise response rates for their courses, and explored instructor and course characteristics as contributing factors. A comprehensive regression model was evaluated to determine the most effective tactics and characteristics. Using incentives had the most impact on response rates. Other effective tactics that increase response rates include reminding students to take the evaluation, explaining how the evaluations would be used to improve instruction, sending personal emails and posting reminders on Blackboard®. Incentives are not widely used; however, findings suggest that non-point incentives work as well as point-based ones, as do simple-to-administer minimum class-wide response rate expectations (compared to individual completion).

Keywords: student evaluation of teaching; response rates; course evaluation; online course evaluation; incentives

Introduction

Course evaluation (often termed student evaluation of teaching or SET) is pervasive in higher education. Research on its effectiveness, limitations and intended uses are well documented. For the last decade, we’ve seen an increasing number of universities moving from pen-and-paper evaluations to SETs delivered in an online environment. Though various reasons are given, the most common include cost savings, ability for students to have more time to complete the SET and increased student anonymity (Ravenscroft and Enyeart 2009).

The literature suggests that, as a norm, response rates for online evaluations are lower than for paper-and-pencil, in-class evaluations by 20–30%. Paper evaluations generally are in the 70–80% range, while online evaluations are in the 50–60% range, but lower response rates do not effect average ratings (Avery et al. 2006; Dommeyer et al. 2004; Guder and Malliaris 2013). Ironically, although SETs are well researched, faculty who regularly conduct research in their own disciplines tend to discount research regarding the reliability, validity and accuracy of online SETs.
with lower response rates (Dommeyer, Baum and Hanna 2002). Nulty (2008) cast
doubt on the ability to determine a statistically reliable and socially acceptable lower
threshold for SET response rates, given the large potential for sample bias in various
sized classes. The purpose of this research was to help faculty identify effective
means, under their control, to increase response rates in their classes to the extent
possible. Increasing response rates is important in order to alleviate concerns about
the generalisability of the evaluation results.

Background and literature review

SETs date back to at least 1915 (Wachtel 1998). The purpose and use of SETs has
expanded from an opportunity to improve teaching to also include a means by
which personnel decisions may be made. This literature review will focus on the
research related to the validity of SETs, on the factors that have been found to
influence ratings and factors that impact response rates in online versions of SETs.

SETs are widely researched. Marsh and Dunkin (1992) claim there have been
thousands of SET validity studies. There is an ongoing debate as to the validity of
SETs with some citing evidence that SETs ratings are positively correlated with stu-
dent learning (Aleamoni and Hexner 1980; McKeachie 1990) and that ratings are
positively correlated with alumni surveys indicating that the effect of time does not
alter the impression of good teaching (Feldman 1989; Overall and Marsh 1980).
Others argue that SETs are inappropriately used as the sole basis to evaluate teach-
ing effectiveness and tenure decisions (Rutland 1990; Zoller 1992), which weakens
their validity since they are not used as intended. Faculty may also view SETs with
hostility (Franklin and Theall 1989) and cynicism (Powell 1978) due to lower
response rates. Nevertheless, despite the faculty questions and sometimes hostility,
SETs are used extensively throughout higher education as a feasible means of
evaluating and/or improving teaching quality.

Numerous studies, most from the era of paper-and-pencil evaluations, have tried
to determine how different characteristics of the course, student or teacher may
relate to variance in the overall Likert-type ratings students fill out. For example,
regarding course characteristics, upper division courses were found to receive higher
ratings than lower division (Feldman 1978; Marsh 1987). This effect is reduced
when other considerations (e.g. class size, expected grade, etc.) are controlled. Like-
wise, class size has mixed findings. Most studies find that smaller classes have
higher ratings (Feldman 1978; Franklin, Theall, and Ludlow 1991). Others have sug-
gested that the instructor’s comfort level with teaching large classes is the determin-
ing factor in whether or not class size effects SET rating (Scott 1977). Still others
find a curvilinear relationship, where large and small classes have higher ratings and
mid-sized classes have lower ratings (Feldman 1984; Koushki and Kuhn 1982).

Just as instructor comfort level with class size may play a role in ratings, other
instructor characteristics have been investigated. Instructor rank/position seems to
play less of a role than teaching experience. First-year teachers are rated lower than
experienced instructors (Centra 1978). Specific rank, however, doesn’t appear to
effect ratings (Feldman 1983) except that professors are rated higher than teaching
assistants (Brandenburg, Slinde, and Batista 1977). Likewise, research is mixed in
the effect of research productivity and ratings. Most studies show no relationship
between research productivity and teaching evaluations (Aleamoni 1999; Feldman
1987; Marsh and Dunkin 1992), although Allen (1996) found a small positive
correlation. Research is also mixed with respect to instructor gender. Some suggest a bias against women (Basow 1994; Koblitiz 1990), while others found a bias for women (Tatro 1995). Feldman’s (1992) meta-analysis found no gender effect.

Evaluation administration has been extensively examined as well. The timing of evaluations was found to have no effect on ratings: that is, ratings remain consistent whether collected at the end of the first week or the last week of classes (Feldman 1979; Frey 1976). However, Aleamoni (1981) and Braskamp, Brandenburg, and Ory (1984) suggest that other timing-related factors can negatively affect student ratings, i.e. just before or after an examination, or during a period of student stress (like final examination week).

More recently, considerable attention has turned to whether or not online evaluations differ from pencil-and-paper format. A number of studies comparing online with pencil-and-paper format have consistently found that ratings remain the same or are slightly improved in the online format (Anderson, Cain, and Bird 2005; Avery et al. 2006; Dommeyer et al. 2004; Ernst 2006; Fraze et al. 2003). Also, qualitative responses tend to increase online (Anderson, Cain, and Bird 2005; Handwerk, Carson, and Blackwell 2000; Hardy 2003; Heath, Lawyer, and Rasmussen 2007). Other advantages of online course evaluations include more class time for teaching and greater student accessibility, as well as reduced administration costs, staff time and results distribution lag (Ballantyne 2003; Bothell and Henderson 2003).

While ratings do not appear to be affected by switching to the online format, response rates for online evaluations consistently decline (Avery et al. 2006; Cody 1999; Dommeyer, Baum, and Hanna 2002; Dommeyer et al. 2004; Ernst 2006; Hmieleski 2000; Hmieleski and Champagne 2000). Nulty (2008) reviewed response rates in nine comparison studies (pencil-and-paper vs. online). All but one study showed that online response rates were much lower than paper-and-pencil evaluations: an average of 23% lower. Faculty have been concerned about the types of students most likely to respond in an online environment. Thorpe (2002) found that the following groups were more likely to respond: women, students who earned the highest grades and students with higher overall GPAs.

There is a broad spectrum of tactics that might be used to increase online SET response rates. On an institution-wide scale, some schools make a student’s early access to grades contingent upon completing his/her course evaluations. Likewise, advertising or lotteries (Nulty 2008) are sometimes used to raise awareness and motivation to complete evaluations.

The logistics of administering SET instruments can also affect response rates. Factors might include how much time is made available to complete the evaluations, or the number and types of reminders sent to students and/or faculty (Nulty 2008). Similarly, attempts are often made to make it easier for students to access the evaluation or reduce the number of short response-type questions to decrease the perceived effort to complete the evaluation. Nulty (2008) analysed five studies that reported using one or more ways to increase response rates as well as those that did nothing. Those who took no measures experienced response rates at or below 23% (Johnson 2002), while institutions that used more measures or incentives received higher response rates.

A further category of tactics to increase SET response rates are those available to the instructor. More benign tactics include efforts by the instructor to personally ask/remind students to complete the evaluations (Norris and Conn 2005), or to discuss why the feedback is important. More active tactics include portraying the SET as an
assignment for the class (Ravenscroft and Enyeart 2009), or using incentives that involve granting extra credit points (Johnson 2002) or non-point rewards, such as bringing a page of notes to the final examination. In light of the low response rates for online evaluations and the concern expressed by academics regarding their validity, Nulty (2008) attempted to define adequate response rates. He argued that sample bias is likely, but probably neither predictable nor randomly distributed. To address these challenges, he calculated a range of response rates for different class sizes, between a best and worst scenario defined by sampling error and confidence-level assumptions. Under the best conditions, the minimum acceptable response rate for a 20-student class was 58%; and, for an 80-student class a 25% rate was acceptable. Given unknown sample bias effects, Nulty concluded that institutions need to raise response rates as far as possible, which means using more and more effective, tactics.

While there is limited research (Crews and Curtis 2011; Dommeyer et al. 2004; Murphy 2004) on the use of instructor-based tactics to increase response rates, this research focused on the size of grade incentive necessary, and on profitable tactics to introduce online course evaluations to a campus or with limited sample size. Little work has been done to sort out the effectiveness of specific tactics in light of their combined use or in terms of how the tactics are configured. We designed this study to provide insight into the effectiveness of various instructor-based tactics and incentives for increasing response rates.

This study was designed to address the following questions:

- What tactics do instructors use to improve their course evaluation response rates? Do instructors using these tactics have higher response rates compared to those who do not?
- Does the number of tactics that instructors employ affect response rates? Do instructors who report using more tactics have higher response rates? Is there a ceiling to the effects of additional tactics?
- Do response rates change based on course and instructor characteristics?
- What is the best model for explaining variation in response rates? How much variation does the model explain?
- Is there an underlying relationship between using tactics and course/faculty characteristics that also needs to be considered?

**Methodology**

This study was conducted at a public university in the north-western United States. Response rate data were from the fall 2012 semester, one year after the university switched to online course evaluations university-wide. For fall 2012, the overall university response rate was 55% across 76,172 student-course enrolments. The evaluation included 19,399 students, taking 3213 separate courses, taught by 1134 different instructors. Response rates varied widely across campus. For example, one department averaged 82%, while another averaged 31%. Individual courses varied from 0 to 100%.

In addition to course evaluation data, we conducted a survey of instructors teaching fall 2012 courses after the semester concluded. The survey asked questions about tactics faculty used to ‘increase student response rates or otherwise obtain student feedback’, with a follow-up that asked them to further describe the incentives
they employed if they selected that option. The survey also asked faculty for some basic demographic information on their tenure-track status, gender and years of teaching.

The survey was sent to all instructors who taught one or more courses during the term. Of the 1134 requests, 56% (678) of the instructors completed the survey. The responding group was fairly representative of the entire instructor population, although those responding to the survey had a somewhat higher response rate (61% vs. 55%) in their courses and included rather more full-time faculty (tenure/tenure-track, lecturers) than the general population (66% vs. 57%). This group of respondents was more evenly balanced in terms of gender, with slightly more women (53% vs. 58% overall). A majority (60%) had seven or more years of teaching experience.

Two-thirds of instructors taught more than one-course section during fall 2012. For those who taught two or more classes, two-course sections were randomly selected for the instructor to respond in the survey: thus, while 678 individuals completed the survey, the data-set was comprised of 1128 classes for which tactics to encourage higher response rates were described. This amounted to nearly 35% of all course sections taught that term. The courses included in the study were also representative of the university as a whole: more were undergraduate courses (87% vs. 80%), 88% (vs. 86%) were taught face-to-face, 11% (vs. 12%) were taught as internet courses and 87% (vs. 88%) enrolled 40 or fewer students.

The survey data were merged with the course evaluation data to form the data-set used in the analysis. The analysis used t-tests to answer the simple set of questions as to whether faculty who reported using a given tactic had higher response rates than those who did not use the tactic. One-way analysis of variance tests were used to check for differences in response rates based on number of tactics used, as well as course and instructor characteristics.

In order to develop a model to explain the variability in response rates, all variables which reached statistical significance using \( p \leq 0.05 \) (including tactics, course characteristics and instructor characteristics variables) were included in a stepwise regression. In this case, the significance level was set at 0.15 for both selection into and removal from the model. A final regression model used the set of variables from the stepwise regression along with all interactions with significant class and instructor variables.

**Results**

*Tactics used by instructors and differences in response rates*

Clearly, employing tactics to improve response rates had an effect. The instructors who indicated they did nothing (13%) averaged a 50% response rate, compared to the 63% average for those who did something. Table 1 shows that two tactics were used more than twice as often as any other: reminding students in class to complete their evaluations (61% of courses), and explaining how the evaluation results would be used to improve instruction (57%). In addition to being widely used, these tactics were fairly effective, associated with fairly large increases in response rates, around 8%.

All of the tactics except two – using the formal mid-semester assessment process (administered by the Centre for Teaching and Learning) and using classroom assessment techniques to obtain feedback throughout the semester – showed significant,
positive impacts on response rates. While the two most widely used tactics were fairly effective, a far lesser used tactic – incentives – produced an average response rate 22% higher than in its absence.

**Follow-up on the use of incentives**

While only one of five (22%) instructors offered some type of incentive to complete the evaluation, the average response rate in classes with incentives was 79% compared to 57% in those without. The prominent impact of incentives demanded further investigation.

The use of incentives has been hotly debated at the institution, so follow-up questions on the survey asked about the types of incentives used. Incentives can be broadly grouped into two larger categories: point-based vs. non-point based, and individual vs. class-wide. For example, an individual/non-point-based incentive might be ‘if the student completes the evaluation, he or she can bring a 3 × 5 card of notes into the final examination’. Or, a class-wide/point-based incentive could be ‘if 80% of the class completes the evaluation, then everyone gets five points added to their final grade’. Point-based incentives were by far the most common, used by over three-quarters of the instructors employing incentives. Points ranged from a low of 0.25 to 1% of the total course grade, with the median closer to 0.5%.

For the subset of 251 instructors who responded that they used incentives, a model to explain response rates was developed using two independent variables: (1) whether a point or non-point incentive was used, and (2) whether receipt of an incentive was based on the individual completing the survey, or whether it was based on a proportion of the entire class completing the survey. Results indicated
that the model did not significantly explain differences in response rates, $F(2248) = 0.32, p = 0.73$. It appears, therefore, that point- and non-point-based incentives work equally well, as do class- and individual-based incentives. Because class-based incentives require far less work to administer, instructors who use incentives should be encouraged to employ the class-based approach because of its efficiency.

A quite unexpected observation involved how to make class-based incentives effective. When describing their use of class-based incentives, many respondents identified the threshold they set for the class to receive its point or non-point reward. The response rate thresholds ranged from 60 to 100%, as illustrated in Table 2. When a lower threshold was set at 60 or 70% of the class completing their evaluations, response rates averaged 62 and 68%, respectively. When higher thresholds were set, at 80, 90 or 100%, higher average response rates resulted of 86, 87 and 89%. No statistical tests were applied to these data and the number of respondents was small, but the pattern was intriguing and deserves consideration.

**Number of tactics used and response rates**

Did the number of tactics that an instructor reported using relate to response rates? As shown by Table 3, only 13% of instructors reported using no tactics to improve response rates, and 70% used at least two tactics. Using a one-way ANOVA, we found that response rates varied significantly based on the number of tactics employed, $F(7,1119) = 23.73, p < 0.0001$. Following with a Tukey’s range test for post hoc comparisons, we found that instructors who employed one tactic had response rates that were similar to those who employed no tactics. Those who employed four or more tactics had significantly higher response rates than any of the other response-rate groups. It appears, therefore, that instructors should try a number of approaches if they are serious about wanting to improve their response rates.

**Class and instructor characteristics and response rates**

As noted in the literature review, a number of characteristics can potentially influence response rates in certain courses. For this study, we looked at the course characteristics of class size, class level and instruction mode. In addition, we included information on instructor type (tenured/tenure-track vs. adjunct) and instructor gender.

As shown by Table 4, all characteristics showed statistically significant differences in response rates with the exception of gender. Small classes had higher response rates than either medium or large classes (64% vs. 61% and 59%).

### Table 2. Impact of threshold levels for incentives.

<table>
<thead>
<tr>
<th>Threshold response</th>
<th>% of faculty with incentives who used threshold (and number)</th>
<th>Average % response</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>11 (8)</td>
<td>89</td>
</tr>
<tr>
<td>90% +</td>
<td>14 (10)</td>
<td>87</td>
</tr>
<tr>
<td>80% +</td>
<td>31 (22)</td>
<td>86</td>
</tr>
<tr>
<td>Sliding scale</td>
<td>11 (8)</td>
<td>73</td>
</tr>
<tr>
<td>70% +</td>
<td>17 (12)</td>
<td>68</td>
</tr>
<tr>
<td>60% +</td>
<td>4 (3)</td>
<td>62</td>
</tr>
<tr>
<td>Unknown</td>
<td>13 (9)</td>
<td>74</td>
</tr>
</tbody>
</table>
Response rates for online courses (50%) were below those of in-person (63%) or hybrid (66%) courses. Graduate courses, and courses taught by tenure/tenure-track faculty, also had higher response rates than comparative groups.

**Best model to predict response rates based on tactics, course and faculty characteristics**

Individually, a number of tactics and characteristics show differences in response rates. But as a group, what is the best set to model response rates? Also, how much variability can the model explain? In order to answer this question, we ran a step-wise regression model using SAS that included all of the previous variables that exhibited a statistically significant difference in response rates. The resulting model predicted 31% of the variability in response rates and was statistically significant, \(F(7,1119) = 72.05, p < 0.0001\).

The most important variable in predicting response rates was whether the instructor used incentives or not. This one tactic explained two-thirds (21%) of the variability in response rates. After that, the number of tactics employed by the instructor improved the model by another 3%. Smaller improvements were contributed by whether the instructor was an adjunct, if the course was taught online, and the course size and level. The final variable added to the model was whether or not the instructor took class time to have students complete their evaluations.

<table>
<thead>
<tr>
<th>Number of tactics</th>
<th>N in group (% of total)</th>
<th>Response rate (%)</th>
<th>Post-hoc group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>149 (13)</td>
<td>50</td>
<td>0 and 1 are similar, lower than all other groups</td>
</tr>
<tr>
<td>1</td>
<td>188 (17)</td>
<td>54</td>
<td>0 and 1 are similar, lower than all other groups</td>
</tr>
<tr>
<td>2</td>
<td>271 (24)</td>
<td>60</td>
<td>2 and 3 are similar, lower than 4-5, higher than 0-1</td>
</tr>
<tr>
<td>3</td>
<td>262 (23)</td>
<td>64</td>
<td>2 and 3 are similar, lower than 4-5, higher than 0-1</td>
</tr>
<tr>
<td>4</td>
<td>155 (14)</td>
<td>71</td>
<td>4 and 5 are similar, all other groups are lower</td>
</tr>
<tr>
<td>5 or more</td>
<td>102 (9)</td>
<td>74</td>
<td>4 and 5 are similar, all other groups are lower</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group members</th>
<th>N in group</th>
<th>Response rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class size</td>
<td>Small: 20 or less</td>
<td>477</td>
<td>64</td>
</tr>
<tr>
<td>(F(2,1124) = 4.87, p = 0.0078)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium: 21–40</td>
<td>496</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Large: &gt;40</td>
<td>154</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Instruction mode</td>
<td>Hybrid</td>
<td>40</td>
<td>66</td>
</tr>
<tr>
<td>(F(2,1111) = 21.82, p &lt; 0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-person</td>
<td>955</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>119</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Course level</td>
<td>Graduate</td>
<td>153</td>
<td>65</td>
</tr>
<tr>
<td>(F(1,1125) = 5.27, p = 0.0218)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>974</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Faculty gender</td>
<td>Female</td>
<td>573</td>
<td>62</td>
</tr>
<tr>
<td>(F(2,1116) = 0.37, p = 0.6877)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>510</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Not recorded</td>
<td>36</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Faculty type</td>
<td>Full-time faculty</td>
<td>748</td>
<td>64</td>
</tr>
<tr>
<td>(F(1,1125) = 25.78, p &lt; 0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-time faculty</td>
<td>379</td>
<td>57</td>
<td></td>
</tr>
</tbody>
</table>
Based on the order of entry of the variables, we can conclude that what the instructor does to improve response rates has a greater effect on response rates than course and instructor characteristics. However, the type of faculty member, instruction mode, class size and level all can and do influence response rates and should be accounted for in response rate comparisons.

**Relationship between tactics and course/faculty characteristics: a final model**

The model above included three variables related to tactics that instructors use: use of incentives, number of tactics and using class time to take the evaluation. It also included one instructor characteristic (adjunct vs. tenure/tenure-track) and three course characteristics: mode of instruction, class size and undergraduate vs. graduate course level. Is there a relationship between the tactics employed and the instructor and course characteristics, i.e. do interactions exist between the tactics employed and the characteristics of the course and instructor?

To address this question, 12 interaction terms were added to the model: four were the interactions between offering incentives, whether the instructor was an adjunct, and whether the course was small, taught online and offered at the undergraduate level. Another four interaction terms were between the number of tactics and the instructor and course characteristics. The final four interaction terms were based on offering time during class to take the evaluation and the instructor and course characteristics.

Table 5 displays the results of the final model. Only one significant interaction met the $p < 0.15$ condition to remain in the model – the interaction between number of tactics employed and course level (graduate or undergraduate). The model as a whole was statistically significant, $F(8,1118) = 65.77$, $p < 0.0001$, with $R^2 = 0.32$. Interpreting the size of the standardised $\beta$ coefficients, we see that the most important factors in the model were the number of tactics employed and the use of incentives to improve response rates. Courses that were taught by adjuncts as well as those taught online had lower response rates, as indicated by the negative sign for the parameter estimates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate ($b$)</th>
<th>Standardised estimate ($\beta$)</th>
<th>Standard error</th>
<th>t-value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.49</td>
<td>0</td>
<td>0.03</td>
<td>18.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Provided incentive</td>
<td>0.20</td>
<td>0.41</td>
<td>0.01</td>
<td>14.87</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Number of tactics used</td>
<td>0.06</td>
<td>0.43</td>
<td>0.01</td>
<td>6.10</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Took class time to complete evaluation</td>
<td>0.07</td>
<td>0.07</td>
<td>0.02</td>
<td>2.91</td>
<td>0.0037</td>
</tr>
<tr>
<td>Small class (&lt;=20)</td>
<td>0.04</td>
<td>0.09</td>
<td>0.01</td>
<td>3.17</td>
<td>0.0016</td>
</tr>
<tr>
<td>Taught online</td>
<td>-0.09</td>
<td>-0.14</td>
<td>0.02</td>
<td>-5.64</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Undergraduate course</td>
<td>0.04</td>
<td>0.06</td>
<td>0.03</td>
<td>1.31</td>
<td>0.1890</td>
</tr>
<tr>
<td>Taught by adjunct</td>
<td>-0.05</td>
<td>-0.12</td>
<td>0.01</td>
<td>-4.60</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Interaction: N of tactics and undergraduate course</td>
<td>-0.04</td>
<td>-0.32</td>
<td>0.01</td>
<td>-3.92</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Note: $F(8,1118) = 65.77$, $p < 0.0001$, $R^2 = 0.32$. 

966 J. Goodman et al.
The interaction between number of tactics employed and course level was also a strong factor in the model. Figure 1 displays the response rates based on the interaction between the number of tactics and whether the course is at the undergraduate or graduate level, holding all other factors at zero. Note that, after the first tactic, the incremental effect on response rate of using additional tactics has a stronger impact on graduate than on undergraduate courses.

**Summary and conclusions**

The purpose of this study was to help identify effective means that faculty could use to increase response rates to online course evaluations in their classes. Results showed that providing incentives to take the evaluation, reminding students in class to take the evaluation, taking class time to complete the evaluation, explaining how the evaluations would be used to improve instruction, sending personal emails, posting reminders on Blackboard® and using some other type of mid-semester check-in with students were all helpful in improving response rates.

Of all these approaches, the use of incentives was the single most effective tactic to improve response rates, by a wide margin. Nevertheless, only a minority of instructors (22%) use them, perhaps due to faculty apprehensions over incentive use to encourage evaluation responses. This study’s findings may help to address two potential concerns of faculty. First, we found that an instructor does not need to use a point-based incentive if concerned that it might directly affect student grades; non-point approaches (such as permitting students to bring a note card to the final examination, making the final optional or even providing some treats for class) work equally well to raise response rates. The second finding was that a simple
class-based approach – setting a high threshold response rate for the whole class to meet – was as effective as the far more time-intensive effort to monitor and reward individual student completion. Further research into possible ratings effects is needed. More research is needed to ensure that incentives do not influence evaluation ratings assigned by students, given anecdotal comments by some instructors who raised such bias as a deterrent to using incentives.

Whether an instructor chooses to use incentives or not, this study also found that the more tactics applied the better, especially for graduate courses. The first tactic used adds an average of 5% response. Each additional tactic (up to five) increases response rates by an extra 5% for graduate courses, or 2–3% for undergraduate courses, on average.

Finally, one of the most important remaining questions is that surrounding the low response rates of students in online courses. This study found that online courses averaged 50% response, compared to 63% in face-to-face courses. This gap suggests that there may be other factors influencing response rates in online classes, perhaps calling for different tactics. As instruction continues to shift to more online coursework, the issue of evaluating the quality grows in importance.

Notes on contributors
James Goodman is the associate director of the Center for Teaching and Learning and an associate professor in the Music Department at Boise State University. His research interests include use of evidence-based best practices to improve teaching effectiveness.

Robert Anson is a professor in the Information Technology and Supply Chain Management Department at Boise State University. His research interests include IT support practices, electronic meetings and group support systems with an emphasis on facilitation in these settings.

Marcia Belcheir is the associate director of Institutional Research at Boise State University. Her main research interest is in the area of student success including factors related to learning, retention and graduation.

References


