Fudging the Numbers: Distributing Chocolate Influences Student Evaluations of an Undergraduate Course

Robert J. Youmans and Benjamin D. Jee
University of Illinois at Chicago

Student evaluations provide important information about teaching effectiveness. Research has shown that student evaluations can be mediated by unintended aspects of a course. In this study, we examined whether an event unrelated to a course would increase student evaluations. Six discussion sections completed course evaluations administered by an independent experimenter. The experimenter offered chocolate to 3 sections before they completed the evaluations. Overall, students offered chocolate gave more positive evaluations than students not offered chocolate. This result highlights the need to standardize evaluation procedures to control for the influence of external factors on student evaluations.

Student evaluations of courses and instructors are commonplace at American universities. These evaluations are an important source of information for instructors seeking to improve their teaching methods, for academic departments that wish to evaluate instructors’ teaching effectiveness, and for students making decisions about enrolling in a course. The influence of student evaluations raises one obvious question: Are they a reliable and valid source of information about course quality and instructor effectiveness?

Compared to other proposed rating systems (e.g., peer review), student evaluations might have superior reliability and external validity (Cashin, 1995). Nevertheless, evaluations can be systematically biased by a number of factors. We delineate these into two main sources of influence, internal and external factors. Much of the interest in the reliability and validity of student evaluations has centered on internal factors, factors that are more or less inherent to a course. For example, prior work has found that student evaluations are related to grading leniency (Greenwald & Gilmore, 1997) and the gender of the instructor (Bassow & Silberg, 1987). Other internal factors might include the relative difficulty of the course, the availability of course materials, the presentation of course materials, and so on. Because internal factors appear to be inextricably linked to an individual course, it is quite difficult to standardize them across courses.

It is also possible that external factors, factors that are entirely independent of the course, also influence student evaluations. Such factors are particular to the evaluation context and could include the timing, format, and distribution of the evaluations, to name a few. Although the influence of external factors might seem intuitive (e.g., many instructors would not want to be evaluated after passing back an exam where students performed poorly), there is little experimental evidence about their effects. This lack of evidence is somewhat surprising, because, compared to internal factors, external ones are far easier to control and manipulate.

In this study, we address the effects of a single but potentially widespread external influence on students’ evaluations: whether students are offered candy before completing their evaluations. We selected this external factor because it is both easily manipulated and based on an informal polling of colleagues, it is a fairly common practice when collecting evaluations. In an attempt to control for potentially influential internal factors, this study included students who were enrolled in the same lecture section of a course, but who attended separate discussion sections. These sections met weekly on the same day and were led by the same teaching assistant. We also attempted to control for confounding external variables: Each discussion section had a maximum enrollment of 20 students, each section met on Friday mornings, all students received the same amount of time to complete their evaluations, and the same experimenter administered all evaluations. The experimental manipulation involved offering students in half of the discussion sections a piece of chocolate prior to completing their evaluations. The students in the other sections were not offered chocolate. If this positive external event affects students’ evaluations, then students who are
offered chocolate might rate the course and its instructor more favorably.

Method

Participants

The participants were 98 undergraduates from the University of Illinois at Chicago from three different classes, two statistics classes (n = 34 and 29) and one research methods class (n = 35). The same instructor taught all three classes. Each class required students to enroll in one of two Friday discussion sections of approximately equal size. The same teaching assistant led sections for each class (with different teaching assistants serving each of the three classes).

Procedure

During the ninth week of instruction, all participants completed an informal midsemester evaluation about the lecture section of their course (the discussion sections were not evaluated at this time). An experimenter who was not involved with the course administered the evaluations at the beginning of class. Students required approximately 10 min to complete evaluations. All students received the same nine-question form (see Table 1). For each question, the student provided a rating from 1 (very poor) to 5 (excellent). The questions pertained to various dimensions of the course and instructor and were adapted from questions appearing in the standard end-of-semester evaluation form used by the University of Illinois at Chicago.

Results

The results were analyzed in a 2 (condition: chocolate vs. control) × 3 (class: statistics 1 vs. statistics 2 vs. research methods) × 9 (evaluation form question: 1 to 9) ANOVA, with condition and class as between-subject variables and evaluation form question as a within-subjects variable. This analysis revealed a main effect of class, \( F(2, 92) = 5.60, p < .05 \); a main effect of question, \( F(8, 92) = 16.02, p < .05 \); and an interaction between class and question, \( F(2, 92) = 2.19, p < .05 \). These effects are tangential to the main question of this study and are not discussed further. The question of main interest is whether the participants offered chocolate before

Table 1. Questions Included in the Evaluation Form and Mean Responses in the Chocolate and Control Conditions

<table>
<thead>
<tr>
<th>Question</th>
<th>Chocolate M</th>
<th>Chocolate SD</th>
<th>Control M</th>
<th>Control SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. You are finding this course intellectually challenging and stimulating.</td>
<td>4.04</td>
<td>0.76</td>
<td>3.63</td>
<td>0.76</td>
</tr>
<tr>
<td>2. Instructor is enthusiastic about conducting the course.</td>
<td>3.92</td>
<td>0.88</td>
<td>3.58</td>
<td>0.90</td>
</tr>
<tr>
<td>3. Course materials are well prepared.</td>
<td>4.50</td>
<td>0.79</td>
<td>4.48</td>
<td>0.77</td>
</tr>
<tr>
<td>4. Students are encouraged to participate in class discussions.</td>
<td>3.78</td>
<td>0.93</td>
<td>3.44</td>
<td>0.90</td>
</tr>
<tr>
<td>5. Instructor's explanations are clear.</td>
<td>4.02</td>
<td>0.94</td>
<td>4.00</td>
<td>1.03</td>
</tr>
<tr>
<td>6. Instructor is friendly toward individual students.</td>
<td>4.20</td>
<td>0.88</td>
<td>3.90</td>
<td>0.99</td>
</tr>
<tr>
<td>7. Methods of evaluating student work were fair and appropriate.</td>
<td>4.08</td>
<td>0.97</td>
<td>4.10</td>
<td>0.88</td>
</tr>
<tr>
<td>8. Compared with other courses you have had at this university, this course is:</td>
<td>3.98</td>
<td>0.85</td>
<td>3.60</td>
<td>0.84</td>
</tr>
<tr>
<td>9. Compared with other instructors you have had at this university, this instructor is:</td>
<td>4.12</td>
<td>0.92</td>
<td>3.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Average rating</td>
<td>4.07*</td>
<td>0.88</td>
<td>3.85*</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Note. Ratings are on a scale from 1 (very poor) to 5 (excellent).

*Means are significantly different at \( p = .05 \).
completing the evaluations gave higher ratings than the participants not offered chocolate. Indeed, this was the case: Participants who were offered chocolate gave higher ratings on average ($M = 4.07, SD = .88$) than participants who were not offered chocolate ($M = 3.85, SD = .89$), $F(1, 92) = 3.85, p = .05$, $d = 0.33$. By Cohen's (1992) standards, this result amounts to a small to medium effect. Moreover, participants in the chocolate condition tended to give higher ratings on almost all of the questions (see Table 1). To address concerns that our findings were due to class standing, we compared final grades in an independent samples $t$ test, which revealed no significant difference in final grades between the chocolate and no chocolate conditions, $t(124) = 1.21, ns$.

Discussion

Given the importance that students, instructors, and departments place on student evaluations, it might be surprising that something as trivial as offering chocolate could affect them. It is beyond the scope of this article to identify how offering chocolate influenced evaluations. Perhaps students perceived the chocolate as a reward, which might have influenced their judgment processes (see, e.g., Arkes, Dawes, & Christiansen, 1986). Perhaps the chocolate put students in a better mood, which had a direct effect on their evaluations (e.g., Schwarz, 2000), or an indirect effect by biasing them to remember mood-congruent information (e.g., Mayer, McCormick, & Strong, 1995; but see also Martin & Clore, 2001). Finally, there is the possibility that experimenter bias contributed to our results, despite the attempts to control for this confound. At this point we cannot be certain of the cause of the chocolate-related effect and, indeed, different students might have responded differently to the offer of chocolate.

Regardless of how the results of this study are explained, the finding that an external factor affected student evaluations motivates several recommendations for both instructors and departments. First, we suggest taking reasonable measures to standardize these factors across evaluation contexts. In our experience, there is wide variety in the time, date, and method for administering evaluation forms, all of which might influence students' evaluations. Although our study focused on a positive external factor, instructors might be more concerned that they are being “shortchanged” simply because of how and when the evaluation forms were administered. Second, we recommend that tenure and awards committees consider the trends of evaluations, rather than any particular sample of performance. This practice could minimize the impact of a single biased set of evaluations. Third, we recommend that department members discuss the collection and use of student evaluations. Some external factors might be more or less universal, but others could be specific to particular universities, departments, courses, and even instructors.

Evaluations are not like a box of chocolates, where instructors never know what they are going to get. Although there is no perfect index of instructor performance, student evaluations remain superior to many other proposed measures (Cashin, 1995). Our findings suggest that standardizing the evaluation context itself could further enhance their reliability and validity.

References


Notes

1. We thank Edward Sargis for allowing us to use his classes for our research.

2. Send correspondence to Robert J. Youmans, Psychology Department, 376 Sierra Hall CSU Northridge, 18111 Nordhoff Street, Northridge, CA 91330–8255; e-mail: robert.youmans@csun.edu.