

## An Elephant in the Room: Bias in Evaluating a Required Quantitative Methods Course

JOSEPH F. FLETCHER  
MICHAEL A. PAINTER-MAIN

University of Toronto

*Undergraduate Political Science programs often require students to take a quantitative research methods course. Such courses are typically among the most poorly rated. This can be due, in part, to the way in which courses are evaluated. Students are generally asked to provide an overall rating, which, in turn, is widely used by students, faculty, and administrators to assess a course. Unfortunately, even questions composed with the best of intentions have the potential to bias the results. In this article, we evaluate the global rating question used at our university and show that it introduces bias into the measure by cueing extraneous considerations. It artificially inflates the number of negative reactions to the course by leading students to think about its required status and their initial level of enthusiasm rather than their level of accomplishment and its value as a learning experience. By locating our results in the course evaluation and framing literature, we suggest an approach to evaluating overall rating questions that can be adapted for use at other institutions.*

**Keywords** course evaluations, framing, quantitative methods

### Introduction

Course evaluations are a common part of contemporary university and college life (Gravestock and Gregor-Greenleaf 2008). Near the end of each term students are typically asked to evaluate their courses on standardized forms. To reduce bias in responses, instructors are asked to leave the room while students complete their evaluations. Although the questions on the forms vary widely across institutions, nearly every evaluation includes at least one question asking students for a general rating of the course or instructor (Gravestock and Greenleaf-Gregor 2008, 27). While a great deal of information is collected on evaluation forms, in practice, students, professors, and administrators tend to focus upon the overall ratings, which are often highlighted when the results are published, ostensibly to aid students in the next term's course selections. Moreover, departmental chairs often refer to these summary ratings in their annual evaluation of the teaching staff (Beran, Violato, and Kline 2007). Promotion and tenure committees do likewise (Beran, Violato, and Kline 2007). Accordingly, professors are also highly interested in this figure.

At our institution the overall rating item is colloquially referred to as the “retake rate.” It is the percentage of students indicating a willingness to retake the class in

Address correspondence to Joseph F. Fletcher, Department of Political Science, University of Toronto, Sidney Smith Hall, Room 3018, 100 St. George St., Toronto, Ontario M5S 3G3, Canada. E-mail: josephf@chass.utoronto.ca

light of their experience in the course. Elsewhere, other overall rating items are commonly in use. As it turns out, there is no consensus as to the validity of these ratings (Gravestock and Greenleaf-Gregor 2008, 32; Drysdale 2010, Ch. 2). In fact, there is “little discussion in the current literature regarding the particular phrasing of global questions” (Gravestock and Greenleaf-Gregor 2008, 32). In an effort to move toward such a discussion, we use a multivariate approach to discern the considerations upon which our students’ overall ratings are based. We trust that this approach can be adapted to assess alternative overall assessment rating questions used at other institutions.

It is not surprising that retake results fluctuate between courses. However, there appears to be a systematic divide between required quantitative-based courses and all others (Buchler 2009; DeVaney 2010; Francis 2011) with quantitative methods courses consistently receiving some of the lowest retake ratings. In this light, we examine closely the overall rating question that appears on the standardized form used in our department and across the Faculty of Arts and Science at our university. It reads “*Considering your experience with this course, and disregarding your need for it to meet program or degree requirements, would you still have taken this course?*” At first glance the question appears to have been constructed, likely with the best of intentions, in at least partial awareness that required courses are often regarded somewhat differently by students than are elective ones, as well as an attempt to have students focus on their course involvement. No doubt the authors of this question were seeking to allow for these facts by inserting the parenthetical phrase “disregarding your need for it to meet degree or program requirements” and the introductory “[c]onsidering your experience with this course.” Presumably, it was thought these instructions would promote an evaluation based on the course itself and would aid students in disregarding the required status of the course in making their evaluations. As we find, even the best of intentions can go awry and have unintended consequences.

## Literature Review

The considerable literature on course evaluations has focused on three overarching themes: the teacher, the student, and the course. Perhaps the most studied aspect of course evaluations is devoted to student evaluations of teaching. In addition to overall teacher ratings, more specific aspects investigated include the implements used in the course by the teacher, as well as a teacher’s personality and behavior toward the students (Clayson 1999; Marsh and Dunkin 1992; Simpson and Siguaw 2000). Student-features studied lie largely with personal identity or traits. These include gender (Tatro 1995), the part/full-time status of the student, enthusiasm going into a class, and the length of university experience (Frey, Leonard, and Beatty 1975). Finally, beyond overall course ratings the nature of the course has been discussed, as well as issues of course difficulty and expected grades (Braskamp and Ory 1994; Marsh 1987; Pounder 2007) and whether the class is required or an elective (Darby 2006; Landis and Pirro 1977; Lovell and Haner 1955; Petchers and Chow 1988; Pohlmann 1975; Pounder 2007; Scherr and Scherr 1990; Wachtel 1998). Our particular interest is with overall course ratings but several other of these factors enter into our analysis as well.

Course factors are of particular interest in understanding the relationship between overall ratings for quantitative courses and question wording. Methods

courses, and those specifically focusing on quantitative methods, are increasingly becoming a mandatory feature of an undergraduate education in the social sciences (Bos and Schneider 2009; Turner and Thies 2009). In their survey of American colleges, Turner and Thies (2009) found that 75% of political science departments possessed a required methods course and that three quarters of such courses involved some element of quantitative data analysis. Research focusing on required courses suggests that students can bring prior attitudes toward the class to bear in their course evaluations. For instance, the literature has consistently indicated that required courses induce lower course evaluations (Darby 2006; Gage 1961; Pohlmann 1975; Scherr and Scherr 1990; Watchel 1998). Indeed, Darby (2006, 28) found that required versus elective status possessed “widespread” influence, affecting views concerning course content, the teacher, and how the course was administered. Watchel (1998, 196) suggests that a potential reason for this is that students hold “lower prior subject interest” for required courses. This fits with other research that noted that subject interest significantly influences overall course ratings (Barth 2008).

Prior interest, or enthusiasm, has also been found to be a good predictor of overall evaluation of a course (Barké, Tollefson, and Tracy 1983; Barth 2008; Francis 2011). Perhaps the poster child for such feelings is the required quantitative methods or statistics course, particularly in the social sciences. Many students are reluctant to take such classes (Alvarez 1992) with dread and resentment being common responses (Coleman and Conrad 2007; Heise 1979, 2002). Onwuegbuzie and Wilson (2003), for instance, found that between 60% and 80% of graduate students had experienced anxiety about courses involving statistics. Lewis-Beck (2001) surmises that this is a result of students lacking an inclination for mathematics, as well as possessing little prior experience in the subject matter. Such prior negative attitudes and beliefs greatly affect how a course is evaluated (Gal and Ginsburg 1994). Hence statistics-oriented courses tend to produce poorer student ratings as compared to nonstatistics courses (Colman and Conrad 2007).

How student evaluations measure views about required, quantitative courses can enhance or minimize reservations students have about such classes. Even a passing awareness of the framing literature suggests that different responses can be elicited by different question wordings. As the work of Kahneman (2011) reminds us, questions can highlight different aspects of the same situation and can focus attention accordingly, with considerable impact upon the answers provided. In the current situation, the designed phrasing, no doubt, to get students to discount their feelings about the required status of the course may well prime or increase the salience or availability of those same feelings. Higgins (1996, 134) tells us that “priming refers to procedures that stimulate or activate some stored knowledge.” In the present instance, the mere mention of a requirement may focus attention, such as to enhance the accessibility of feelings about required courses, inadvertently priming the very considerations that the question is presumably meant to avoid. In addition, by having students focus on their “experience with this course,” the retake question directs respondents to consider their specific involvement and the attendant feelings. Faced with similar issues, survey researchers have become particularly attentive to framing effects that may lead to priming (see Chong and Druckman 2007; Druckman 2004; Gamson and Modigliani 1987; Johnson 2011; Kinder 2003; Nelson and Kinder 1996; Zaller 1992). Nevertheless, there has been little research on framing effects on summary course evaluations (Fox 2006, 87). The relatively few studies

that do exist, while clearly preliminary, nevertheless suggest pronounced framing effects on overall course ratings (Fox 2006; Lynøe, Juth, and Helgesson 2012).

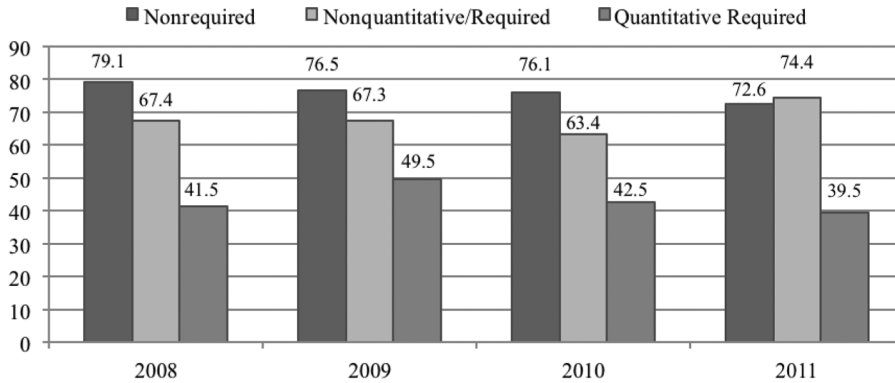
## Methodology

The data for this study came from University of Toronto students taking a full-year quantitative methods course in Political Science between 2009 and 2011.<sup>1</sup> The course is a required class for specialists, a designation applied to those whose primary focus is in the study of politics.<sup>2</sup> In each instance, students voluntarily filled out course evaluations near the end of term.

Insofar as the questions on the standardized course evaluations were not under our control, we could not experiment with their wording. Instead, we took advantage of an opportunity to pose additional questions at the end of the standard evaluation. The additional items appeared on a separate sheet and students were asked to enter their responses on the standardized form in a special area designated for “additional statements or questions which may be supplied in class.” One particular question we asked involved offering an alternative retake question. Consistent with Kahneman’s (2011, 282–288) emphasis on points of reference and his discussion of focusing attention (402–406), we suspected that providing an alternative point of reference would markedly affect the student’s overall evaluation of the course. So we devised an item that sets up a new future reference point rather than focusing attention on the past.

In determining how to construct the question wording for our new retake question we borrowed from the voting literature, which distinguishes between retrospective and prospective evaluations (Downs 1957; Fiorina 1981). As such, we sought to frame a revised course evaluation question in terms of future benefits or utility, that is, prospective evaluations. Hence one of the additional items we included asked the following: “*Considering the value of this course in preparing for future study and future work, would you still have taken this course?*” In selecting this particular wording, we did not mean to arrive at a perfectly neutral question. Instead, we sought to provide students with an alternative frame for the retake item, one that primed prospective considerations rather than the retrospective ones. The choice of focusing on future benefits was thought to be a useful approach, given that one of the selling points of such a course is its ability to give students a skill set in demand both within the academic and work fields. Hill (2002, 113), for example, exhorts the importance of “scientific” education in political science as it plays an invaluable role in modern society and provides students with an ability to conduct social science inquiry outside the immediate setting. Thus, we anticipated a prospective focus might highlight different considerations than the traditional question with its retrospective focus on things past.

Data analyses are conducted using cross-tabulation and logistic regression. Our goal is not to consider all of the potential factors that may affect course evaluations. Rather, our purpose is to determine whether the two versions of the question produce different results and, if so, what key predictors might explain this divergence. For this, we look at factors such as whether the course is a required one for the student, initial enthusiasm, perceived course difficulty, grade expectations, and the value of the course as a learning experience (see the appendix for specific question wording). Our analysis enables us to consider how these factors differentially predict responses to the two forms of the retake question.



**Figure 1.** 2008–2011 Retake rates: Political science course comparison. The raw data were provided by and used with the permission of the Arts and Sciences Student Union (ASSU) and the Department of Political Science at the University of Toronto.

**Findings**

One expectation from the literature is that required, quantitative courses are the least well regarded by students. Initial evidence supports this claim. As shown in Figure 1, as assessed by the tradition question, retake rates were the lowest for the required quantitative course in the political science department under analysis.<sup>3</sup> Across the 2008–2009 and 2011–2012 calendar years, a substantial difference in retake rate persisted between the required second-year quantitative research methods course and other nonquantitative required courses. And the difference between required quantitative courses and nonrequired nonquantitative courses is, for the most part, greater still.<sup>4</sup> These differences suggest that not only is the retake rate for required courses usually lower than nonrequired courses but that there is something distinctive about the quantitative course that breeds lower rates of retake. One potential reason is that students are hesitant to embrace a course involving mathematics. By asking students to focus on their “experience” with the course, the retake question potentially enhances the anxiety that students face with statistics.

Does the standardized retake question promote a particular framing effect? Table 1 presents the results for the traditional retake question as well as for the revised version we introduced. As the left-hand column shows the traditional question produces roughly a 40–60 split with most students indicating that they were unwilling to retake the course. In the right-hand column, however, we see a complete reversal of this distribution. Asking the revised question results in 60% of the respondents in favor of retaking the course.<sup>5</sup> The chi-square analysis indicates

**Table 1.** Willingness to retake required methods course by question type

Retake course?	Traditional question	Revised question
Yes	42.8%	59.4%
No	57.2%	40.6%
<i>n</i>	152	155

Note:  $\chi^2 = 8.46$  with one *df*;  $p = .004$ .

that there is an extremely small chance ( $p < .004$ ) that there is no difference between the question versions. Or more formally, the data-generating process that produced the answers to one question is unlikely to have also produced the answers to the other.

Evidently how one frames the retake question makes a difference. And this difference is not merely statistical. Depending upon which question is used the majority reverses from negative to positive. In particular, framing the retake question in the traditional way produces a majority who say they are unwilling to retake the course. Alternatively, framing the retake question in a way that highlights future considerations produces a majority saying they are willing to retake the course. The practical implications of this are considerable in so far as students, professors, and administrators are likely to draw quite a different conclusion about the course based upon which question wording is used.

Since both the traditional and the revised questions were asked of all respondents, we are able to cross-tabulate student answers to the two questions. The results are presented in Table 2. The resulting pattern is striking. Every student who indicated that he or she would retake the course using the traditional question also indicated a willingness to retake using the revised question. In other words, there were no false positives created by the traditional frame. However, there are a considerable number of false negatives. Over one quarter of those who said they would not retake the course using the traditional question answered that they would retake the very same course using the new question. The chi-square test indicates that there is a strong probability ( $p = .000$ ) that the two columns statistically differ.<sup>6</sup>

Clearly there is a difference between the responses obtained using the two versions of the retake question. But why is there such a difference? What can be said about their respective data-generating processes? In order to investigate, we look at whether different factors predict responses to the two versions of the question. For this we require a multivariate approach. Given the response options are in each instance dichotomous (1 = yes; 0 = no), logistic regression is utilized.

The results are presented in Table 3.<sup>7</sup> In this case, the dependent variable is the probability of answering “yes” rather than “no” to the retake question. Separate equations using the same predictors are presented for both the traditional and revised versions of the retake question.

Looking first at the summary measures presented at the bottom of Table 3, we see that the overall fit ( $R_L^2$ ) of both models is quite good (Menard 2010). The independent variables in the equation markedly assist in explaining variation in the dependent variables. Similarly, the percent of cases correctly classified shows the extent to which respondents are classified into the correct discrete yes and no categories of the dependent variable. For both equations the percentages show

**Table 2.** Response to revised retake question by traditional question response

Revised retake question response	Yes	No
Yes	100%	28.4%
No	0	71.6%
<i>n</i>	63	81

$\chi^2 = 75.6$  with one *df*;  $p = .000$ ; McNemar Exact significance = .000.

**Table 3.** Predictors of traditional and revised retake questions-logistic regression

	Traditional question	Revised question
Precourse Factors		
Required Status	-0.90 (.45)*	-0.58 (.47)
Enthusiasm at Time of Registration	0.70 (.35)*	0.49 (.41)
In-Course Factors		
Level of Difficulty	-0.41 (.25)	-0.47 (.26)
Expected Grade	0.39 (.29)	0.76 (.31)**
Learning Experience	1.05 (.24)***	1.23 (.26)***
Constant	-2.73 (2.76)	-4.59 (2.96)
Summary Measures		
-2 LL (initial; model)	(152.70; 107.50)	(150.90; 96.60)
$R_L^2$ ; Cox & Snell; Nagelkerke	(.30; .33; .45)	(.36; .38; .52)
% of Cases Correctly Classified	75.20	78.60
<i>n</i>	113	112

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

a marked improvement over the roughly 58% modal result found in the frequency distributions of Table 2. So both quantitative and qualitative predictions are robust. The  $R_L^2$  and case classification measures each suggest the fit of the model for the revised question is marginally better than that for the traditional question, but the difference is relatively modest at best.

The main portion of Table 3 presents unstandardized logistic regression coefficients (and their standard errors). Substantive differences exist when examining the relationship between precourse predictors and the two retake questions. For both models, the negative coefficients for required status indicate that, controlling for the other variables in the model, the retake rate is lower among those for whom the course is required. However, viewed within the context of their respective standard errors, the effect of the course being required is substantially greater using the traditional measure than with the revised measure. Moreover, the effect is statistically significant in the traditional model but not in the revised retake model. Similarly, a question asking students about their initial enthusiasm for taking the course significantly predicts the traditional retake results but not the revised ones. We see that an increase in initial enthusiasm is associated with a greater increase in support for the traditional retake compared with the revised question. And viewed within the context of their respective standard errors the effect of students' initial enthusiasm is statistically significant using the traditional question whereas the effect is not beyond what one would expect by chance using the revised question.

Taken together, these findings suggest that precourse factors such as the required status of the course and students enthusiasm going into the class are engaged by the traditional wording used to assess the course retake rate. Of course, this is something we might reasonably expect given how the traditional question highlights both enthusiasm and requirement. What is particularly instructive is that such precourse considerations are not significantly salient when using the revised question, providing clear evidence of a framing effect.<sup>8</sup>

The second section of Table 3 focuses upon in-course factors: students' ratings of course difficulty, the final grade they expect, and their rating of the value of the

course as a learning experience.<sup>9</sup> The course difficulty ratings are not significantly related to either retake question. However, the sign in both instances is as expected; the more difficult the course is thought to be the less likely the student is to indicate a willingness to retake it.<sup>10</sup> The results for expected grades tell a different story. The students' anticipated final grade is not significantly related to the answers given on the traditional retake question. This is in contrast to the revised question where the coefficient is not only positive but statistically significant. This indicates that the revised question engages students' assessment of how well they are likely to do in the class in determining whether they would retake it. This makes sense, of course, because the revised question asks about their future academic and work life for which one's grade could make a difference.

The final item in Table 3 shows the results for the item used by students to assess the value of the course as a learning experience. The more that students see the course as a positive learning experience the more likely they are to say that they will retake it irrespective of which question we use. Even so the coefficients are of different sizes despite having very similar standard errors; the one attached to the revised retake model is approximately 20% larger. So while both forms of the question engage student considerations of the course as a learning experience, the revised question asking about their future does so to a greater extent.

Considered together, the latter three sets of coefficients presented in Table 3 indicate that in-course factors are engaged to a greater extent by the revised course retake question than by the traditional one. Considerations such as how well one is doing in the course and the quality of the course as a learning experience are elicited more by the revised question asking about one's future.

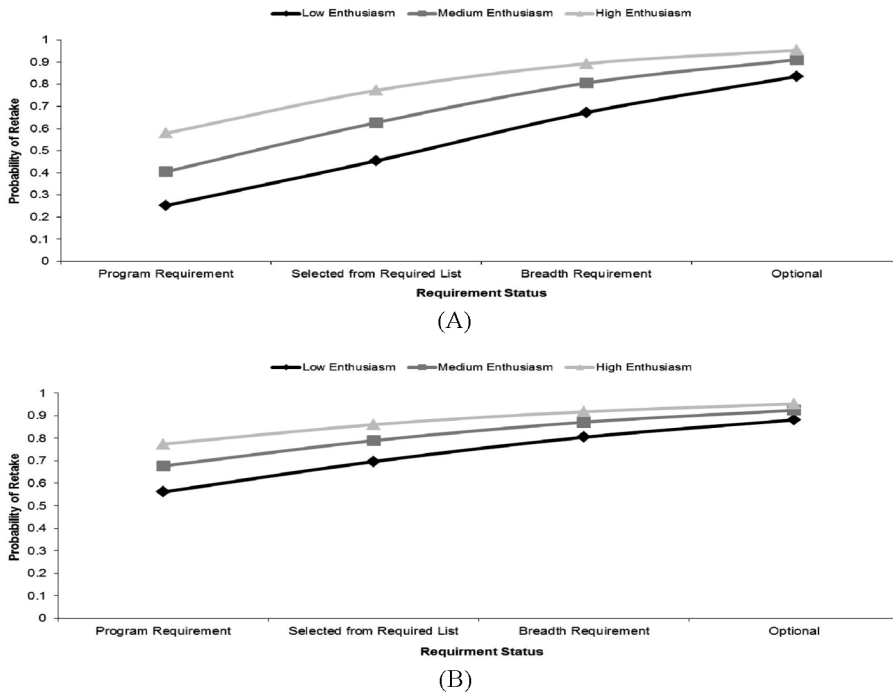
Some aspects of these findings can be better appreciated when viewed in terms of predicted probabilities of students answering yes to each version of the retake question. The two panels of Figure 2 present the results relative to precourse considerations of course status and initial enthusiasm. Those predicted probabilities for in-course factors appear in the panels of Figure 3.<sup>11</sup> In each instance, the upper panel of the figure reports results using the traditional question while the lower panel reports on the revised question.

Looking to the first panel in Figure 2, the influence of requirement status of the course is shown by the upward slope of the predicted probabilities, while initial enthusiasm's impact is evident in the spread of the lines. Consistent with the retrospective focus of the traditional retake question, both precourse factors have a substantial influence on the traditional retake rate. The slopes are markedly upward and the three lines are easily differentiated. Moreover, there is a suggestion that the effect of course status is greater among those with low levels of initial enthusiasm than among those with high levels of enthusiasm.<sup>12</sup>

The comparable probabilities obtained using the revised retake question appears in the lower panel of Figure 2. Here both the slope and spread are substantially muted. Neither status of the course nor initial enthusiasm have much of an effect. And the gap between those with low and high enthusiasm is relatively small irrespective of course status.<sup>13</sup> Quite simply, these precourse considerations are not as engaged by the revised question emphasizing prospective considerations as they are by ones that focus attention on retrospective considerations, as does the traditional one.

The predicted probabilities for in-course factors are presented in Figure 3. Once again, the results for the traditional retake question are reported in the upper panel and those for the revised question appear in the lower. Only in this instance, in-course





**Figure 2.** Probability of retake by precourse factors: (A) traditional retake question and (B) revised retake question.

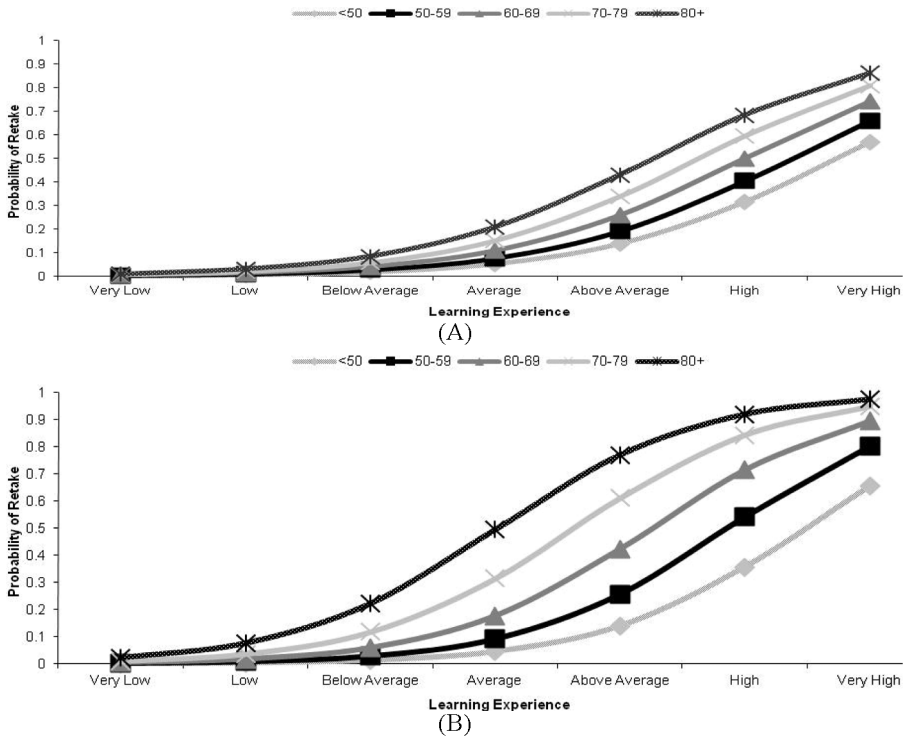
factors, such as anticipated grade and the quality of the learning experience, have relatively less traction with the retrospective focus of the traditional question than with the revised question. The prospective focus of the revised question engages considerations of how well the course provides students with future utility.

Looking at the results as a whole we clearly see different framing effects. The traditional question overestimates the number of students who say that they would not retake the course because the question reminds them of the required status of the course and primes their original level of enthusiasm regarding the class. In contrast, the revised question does not significantly engage these considerations. Instead, students reply to the revised question in terms of their anticipated course grade and the value of the class as a learning experience. These are factors that are likely to be important in future academic and work endeavors. They are evidently cued by the prospective wording of the revised question.

## Discussion

In the opening paragraph of his book entitled *Don't Think of an Elephant*, George Lakoff (2004, 3) tells us that the first thing he does in his Cognitive Science 101 course at Berkeley is to ask his students not to think of an elephant. The point of the exercise is that his students think of an elephant despite his instructing them otherwise.

One of the tenets of cognitive linguistics is that our brains do not automatically process negatives (Kaup, Lüdtkke, and Zwan 2006); questions including negatives



**Figure 3.** Probability of retake by in-course factors: (A) traditional retake question and (B) revised retake question.

must initially be subconsciously processed in the positive in order to fix their meaning. Such is apparently the case with the traditional retake question. In order to parse the meaning of this question asking respondents to disregard the required status of the course, students have had no choice but to make the association and to think of the very concept being linguistically negated. In using the traditional question to assess the retake rate, we are essentially asking students not to think of an elephant. The elephant in the room for most students is that the course is required. And this is what they inadvertently are asked to think about when faced with the traditional retake question. Essentially, it primes thinking about the required status of the course and thereby frames the answers given by the students in just those terms (Kahneman 2011).

The wording of the traditional question was, no doubt, introduced with the idea of eliminating or at least reducing bias against required courses. Unfortunately, good intentions sometimes go awry. Ironically, the traditional question introduces bias into the course evaluation by drawing attention to the required status. Asking students to disregard their need for the course to meet program requirements produces precisely the opposite of what was intended. This is compounded by the fact that students are often wary of taking statistics courses.

Student dread for quantitative courses highlights another priming tool in the traditional retake question. This involves asking students to respond based on their course experience. The current findings provide evidence that this may prime students to evaluate their initial enthusiasm for the course. In this regard, the

traditional retake question is further priming course dissatisfaction, something that may differentiate quantitative courses from other nonquantitative political science courses, be they required or optional.

Again, we do not presume that our reframing of the retake question in terms of future academic and workplace considerations represents an unbiased question. Arguably, it is a better question than the traditional one, but it nevertheless frames the question in a particular way. Perhaps a simpler alternative should be considered such as “*Would you recommend that others take this course?*” Nevertheless, it would be wise to examine responses to this question along the lines we do here in order to determine what sorts of considerations students may draw upon in answering a retake question with a less explicit frame. Similarly, we would urge others to use a similar approach to determine which considerations students rely upon in answering any overall rating items employed in course evaluations at their own institution.

## Appendix

### *Question Wording and Variable Coding*

Question Wording	Item Coding Valid Percent
Traditional Retake Question:	1 (Yes) 42.8
“ <i>Considering your experience with this course, and disregarding your need for it to meet program or degree requirements, would you still have taken this course?</i> ”	0 (No) 57.2
Revised Retake Question:	1 (Yes) 59.4
“ <i>Considering the value of course in preparing for future study and future work, would you still have taken this course?</i> ”	0 (No) 40.6
Required Status:	4 (Program Requirement) 91.9
“ <i>Status of the course for you:</i> ”	3 (Selected from Required List in a Program) 4.1
	2 (Breadth Requirement) 1.4
	1 (Optional) 2.7
Enthusiasm:	1 (Low) 60.5
“ <i>Your level of enthusiasm to take this course at the time of initial registration:</i> ”	2 (Medium) 26.3
	3 (High) 13.2
Level of Difficulty:	1 (Very Low) 0.0
“ <i>Compared to other courses at the same level, the level of difficulty of the material is. . .</i> ”	2 (Low) 1.2
	3 (Below Average) 0.6
	4 (Average) 15.2
	5 (Above Average) 26.2
	6 (High) 34.1
	7 (Very High) 22.5

(Continued)

## Appendix Continued

Question Wording	Item Coding Valid Percent
Expected Grade: “ <i>Your expected grade in this course:</i> ”	1 (<50) 0.6 2 (50–59) 6.4 3 (60–69) 20.5 4 (70–79) 45.5 5 (≥80) 26.9
Learning Experience: “ <i>The value of the overall learning experience is...</i> ”	1 (Very Low) 2.9 2 (Low) 2.1 3 (Below Average) 6.4 4 (Average) 25.0 5 (Above Average) 22.9 6 (High) 25.0 7 (Very High) 15.7

## Notes

1. All classes were taught by the same instructor precluding examination of instructor effects.

2. For example, specialists are required to take half of their 20 credits in political science courses. Political Science majors must take seven. Roughly 90% of those who take the course are required to do so.

3. As part of their requirements, students are, at times, given a pool of courses from which they need to pick one. These courses were treated as required and considered nonquantitative.

4. The 2011–2012 findings suggest that the required, nonquantitative courses eclipsed nonrequired courses in retake rate. One explanation for this is that a popular first-year course became required in 2011–2012.

5. The revised question appears to go some distance toward closing the ratings gaps shown in Figure 1, but any such comparison should be based on similar not different questions.

6. The McNemar test, specifically designed to determine whether the same respondents differ in their answers on two measures, produces an exact significance of .000.

7. The multivariate models exclude approximately 27% of the respondents due to missing data on at least one of the independent variables. Those excluded do not differ significantly from students included on any of the variables in the equation ( $p$  values range from .21 to .99 with an average of .61). Green (1991) suggests a minimum sample size of 50. To increase the sample size responses for 17 cases on the learning experience question were imputed based on a close examination of the answer sheets. Their inclusion does not substantially affect the analysis nor does constraining the analysis to only those required to take the course.

8. Preliminary effort at path analysis suggests that the effect of required status on the traditional retake question may be mediated by initial level of enthusiasm.

9. In selecting items for inclusion in the models, we sought to balance concerns over bias and efficiency. A number of inefficient factors such as questions about the workload, the quality of the readings as well as ratings of the tutorials and laboratory sessions that are part of the course are insignificant as predictors; eliminating them does not substantially affect the model or the conclusions drawn here.

10. This makes sense in that neither the traditional nor revised question evoke considerations of matters such as degree of difficulty, workload, or time commitment entailed by the course. It would be possible, of course, to design a question that does

cue such considerations by asking respondents to consider how much they had to do for the class, how many Saturday afternoons and evenings they had to devote to the course, etc.

11. All other variables in the model are held constant.

12. The gap between them narrows considerably from .33 (.58 - .25) on the left hand side to .11 on the right (.95 - .84).

13. It moves from .21 on the left (.77 - .56) to .70 on the right (.95 - .88).

## References

- Alvarez, Ramon Michael. 1992. "Methods Madness: Graduate Training and the Political Methodology Conferences." *The Political Methodologist* 5(1): 2–3.
- Barké, C. R., Nona Tollefson, and D. B. Tracy. 1983. "Relationship between Course Entry Attitudes and End-Of-Course Ratings." *Journal of Educational Psychology* 75(1): 75–85.
- Barth, Michael M. 2008. "Deciphering Student Evaluations of Teaching: A Factor Analysis Approach." *Journal of Education for Business* 84(1): 40–46.
- Beran, Tanya, Claudio Violato, and Don Kline. 2007. "What's the "Use" of Student Ratings of Instruction for Administrators?" *Canadian Journal of Higher Education* 37(1): 27–43.
- Bos, Angela L. and Monica C. Schneider. 2009. "Stepping Around the Brick Wall: Overcoming Student Obstacles in Methods Courses." *PS: Political Science & Politics* 42(2): 375–383.
- Braskamp, Larry and John Ory. 1994. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco: Jossey-Bass.
- Buchler, Justin. 2009. "Teaching Quantitative Methodology to the Math Averse." *PS: Political Science & Politics* 42(3): 527–530.
- Chong, Dennis and James N. Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10: 103–126.
- Clayson, Dennis E. 1999. "Students' Evaluation of Teaching Effectiveness: Some Implications of Stability." *Journal of Marketing Education* 21(1): 68–75.
- Coleman, Charles and Cynthia Conrad. 2007. "Understanding the Negative Graduate Student Perceptions of Required Statistics and Research Methods Courses: Implications for Programs and Faculty." *Journal of College Teaching & Learning* 4(3): 11–20.
- Darby, Jenny A. 2006. "The Effects of Elective or Required Status of Courses on Student Evaluations." *Journal of Vocational Education & Training* 58(1): 19–29.
- DeVaney, Thomas A. 2010. "Anxiety and Attitude of Graduate Students in On-Campus vs. Online Statistics Courses." *Journal of Statistical Education* 18(1): 1–15.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.
- Druckman, James N. 2004. "Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects." *American Political Science Review* 98: 671–686.
- Drysdale, Michael J. 2010. *Psychometric Properties of Post Secondary Students' Course Evaluations*. PhD diss. Utah State University.
- Fiorina, Morris. 1981. *Retrospective Voting in American Elections*. New Haven, CT: Yale University Press.
- Fox, Craig R. 2006. "The Availability Heuristic in the Classroom." *Judgment and Decision Making* 1(1): 86–90.
- Francis, Clare. 2011. "Student Course Evaluations: Association with Pre-Course Attitudes and Comparison of Business Courses in Social Science and Quantitative Topics." *North American Journal of Psychology* 13(1): 141–154.
- Frey, Peter, Dale Leonard, and William Beatty. 1975. "Student Ratings of Instruction: Validation Research." *American Educational Research Journal* 12(4): 435–447.
- Gage, N. L. 1961. "The Appraisal of College Teaching: An Analysis of Ends and Means." *Journal of Higher Education* 32(1): 17–22.
- Gal, Iddo and Lynda Ginsburg. 1994. "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework." *Journal of Statistics Education* 2(2). <<http://www.amstat.org/PUBLICATIONS/JSE/v2n2/gal.html>> (January 28, 2012).

- Gamson, William A. and Andre Modigliani. 1987. "The Changing Culture of Affirmative Action." In *Research in Political Sociology*, pp. 137–177, Vol. 3, ed. Richard Braumgart. Greenwich, CT: JAI Press Inc..
- Gravestock, Pamela and Emily Gregor-Greenleaf. 2008. *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.
- Green, Samuel B. 1991. "How Many Subjects Does it Take to Do a Regression Analysis?" *Multivariate Behavioral Research* 26(3): 499–510.
- Heise, David R. 1979. *Understanding Events: Affect and the Construction of Social Action*. New York: Cambridge University Press.
- Heise, David R. 2002. "Understanding Social Interaction with Affect Control Theory." In *New Directions in Contemporary Sociological Theory*, eds. Joseph Berger and Morris Zelditch. Boulder, CO: Rowman and Littlefield, 17–41.
- Higgins, E. Tory. 1996. "Knowledge Activation: Accessibility, Applicability, and Salience." In *Social Psychology: Handbook of Basic Principles*, pp. 133–168, eds. E. Tory Higgins and Arie Kruglanski. New York: Guilford Press.
- Hill, Kim Quail. 2002. "The Lamentable State of Science Education in Political Science." *PS: Political Science & Politics* 35(1): 113–116.
- Johnson, Jeffery. 2011. "What Do We Measure?: Methodological Versus Institutional Validity in Student Surveys." Presented at the Association for Institutional Research Annual Forum, Toronto, Ontario.
- Kahneman, Daniel. 2011. *Thinking Fast and Slow*. Toronto: Doubleday Canada.
- Kaup, Barbara, Jana Lüdtkke, and Rolf A. Zwan. 2006. "Processing Negated Sentences with Contradictory Predicates: Is That Door That Is Not Open Mentally Closed?" *Journal of Pragmatics* 38(7): 1033–1050.
- Kinder, Donald R. 2003. "Communication and Politics in the Age of Information." In *Oxford Handbook of Political Psychology*, eds. David O. Sears, Leonie Huddy, and Robert Jervis. New York: Oxford University Press, 357–393.
- Lakoff, George. 2004. *Don't Think of an Elephant: Know Your Values and Frame the Debate*. White River, VT: Chelsea Green.
- Landis, Larry M. and Ellen B. Pirro. 1977. "Required/Elective Student Differences in Course Evaluations." *Teaching Political Science* 4: 405–422.
- Lewis-Beck, Michael S. 2001. "Teaching Undergraduates Methods: Overcoming "Stat" Anxiety." *The Political Methodologist* 10(1): 7–9.
- Lovell, George D. and Charles F. Haner. 1955. "Forced-Choice Applied to College Faculty Rating." *Educational and Psychological Measurement* 15: 291–304.
- Lynöe, Niels E., Niklas Juth, and Gert Helgesson. 2012. "Case Study of a Framing Effect in a Course." *Medical Teacher* 34(1): 68–70.
- Marsh, Herbert W. 1987. "Students' Evaluation of University Teaching: Research Findings, Methodological Issues and Directions for Future Research." *International Journal of Educational Research* 11(3): 253–388.
- Marsh, Herbert W. and Michael J. Dunkin. 1992. "Students' Evaluations of University Teaching: A Multidimensional Perspective." In *Higher Education: Handbook of Theory and Research*, pp. 143–233, Vol. 8, ed. John C. Smart. New York: Agathon Press.
- Menard, Scott. 2010. *Logistic Regression: From Introductory to Advanced Concepts and Applications*. Thousand Oaks, CA: Sage.
- Nelson, Thomas E. and Donald R. Kinder. 1996. "Issue Fames and Group-Centrism in American Public Opinion." *Journal of Politics* 58: 1055–1078.
- Onwuegbuzie, Anthony J. and Vicki A. Wilson. 2003. "Statistics Anxiety: Nature, Etiology, Antecedents, Effects, and Treatments: A Comprehensive Review of the Literature." *Teaching in Higher Education* 8(2): 195–209.
- Petchers, Marcia and Julian Chow. 1988. "Sources of Variation in Students' Evaluations of Instruction in a Graduate Social Work Program." *Journal of Social Work Education* 24(1): 35–42.

- Pohlmann, John T. 1975. "A Multivariate Analysis of Selected Class Characteristics and Student Ratings of Instructions." *Multivariate Behavioral Research* 10(1): 81–91.
- Pounder, James. 2007. "Is Student Evaluation of Teaching Worthwhile?: An Analytical Framework for Answering the Question." *Quality Assurance in Education* 15(2): 178–191.
- Scherr, Frederick and Susan Scherr. 1990. "Bias in Student Evaluations of Teacher Effectiveness." *Journal of Education for Business* 65(8): 356–358.
- Simpson, Penny M. and Judy A. Siguaw. 2000. "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response." *Journal of Marketing Education* 22(3): 199–213.
- Tatro, Clayton N. 1995. "Gender Effects on Student Evaluations of Faculty." *Journal of Research and Development in Education* 28(3): 169–173.
- Turner, Charles C. and Cameron G. Thies. 2009. "What We Mean by Scope and Methods: A Survey of Undergraduate Scope and Methods Courses." *PS: Political Science & Politics* 42(2): 367–373.
- Wachtel, Howard K. 1998. "Student Evaluation of College Teaching Effectiveness: A Brief Review." *Assessment and Evaluation in Higher Education* 23(2): 191–212.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.