

Appendix

Appendix I

Evaluation Questions by Category

Question Type	Question
Instructor	1. Overall the instructor was effective
Instructor	2. Instructor was available for consultation in a timely manner.
Instructor	3. The instructor treated me fairly.
Instructor	4. The instructor treated me with respect
Instructor	5. The instructor welcome and encouraged questions and comments.
Instructor/Course	6. The instructor stimulated student learning
Instructor/Course	7. The instructor presented the information clearly.
Instructor/Course	8. The instructor emphasized the major points and concepts.
Instructor/Course	9. The instructor went beyond presenting the information in the text.
Instructor/Course	10. The instructor demonstrated knowledge of the subject.
Course	11. Overall, this course was a valuable learning experience.
Course	12. The assignments were relevant and useful.
Course	13. Course materials were relevant and useful.
Course	14. Expectations were clearly stated either verbally or in the syllabus.
Course	15. The testing and evaluation procedures were fair.
Course	16. The workload was appropriate for the hours of credit.
Course	17. The textbook or other purchased materials were relevant and useful.
Technology	18. The technology for delivering the instruction was appropriate for the course.
Technology	19. Information was available to help me solve technical problems before they cause delays in my course.
Technology	20. Documentation for accessing electronically delivered class lectures and/or material was effective.
Administrative	21. Registration procedures for the course ran smoothly.
Administrative	22. Advising was available and adequate.
Administrative	23. Course materials were readily accessible.

Appendix II

Comment Themes, Descriptions, and Examples

Theme	Description	Example
Personality	Demeanor, attitude, or other personality characteristics	“Very fun,” “Prickly demeanor,” “Rude”
Appearance	Instructor’s body, clothing, or overall attractiveness	“Best looking teacher at University,” “Nice to look at”
Entertainment	Interesting, engaging, boring, or overall entertaining the instructor or course content was	“Very engaging,” “Boring,” “Very interesting class”
Intelligence/competency	Mastery, qualification, or intelligence of the instructor	“Knows his stuff,” “highly qualified”
Incompetency	Mention or question the instructor’s understanding of or ability to teach the material	“Has no idea what she’s talking about”
Referred to as “professor”	Called the instructor “professor”	“Great professor,” “Professor Mitchell”
Referred to as “teacher”	Called the instructor “teacher”	“Good teacher,” “Rudest teacher”

Appendix III

Content Analysis and Coding

To begin the exploration into gender bias in student evaluations, the authors first examined student *comments* about the instructors. The comments were obtained for all courses, whether online or face-to-face, taught by the instructors in a given semester. Students, at the conclusion of every semester, are asked to complete an evaluation of their instructor and course. Students rate their instructors and courses on several ordinal scale questions, and then are asked to provide comments on the course or instructor.

To ensure inter-coder reliability, each of the authors independently calculated the percentage of comments for each instructor that addressed each theme, and the percentages calculated by each author for each theme were averaged. Comments in official evaluations were

examined separately from those posted on Rate My Professors. To compare comments for Dr. Martin (a man) and Dr. Mitchell (a woman), each percentage for Dr. Mitchell was subtracted from the percentage for Dr. Martin. Thus, a positive difference indicates that Dr. Martin received more comments within the theme than Dr. Mitchell, and a negative difference indicates that Dr. Mitchell received more comments within the theme.

The comments were only used to compare the *type of language* used by the students, *not* whether one instructor was “better” or “worse.” We chose not to include positive or negative commentary because the quality of the instructor is a rather subjective measure: making a direct comparison of the evaluations is difficult, particularly in face-to-face courses with different subject matter. For example, Dr. Martin, a male instructor, taught both a survey of Comparative Politics and a Western European Politics course, while Dr. Mitchell, a female instructor, taught Game Theory, Political Analysis (statistics and research methods), and an Honors section of International Relations. Not only are the difficulty levels of the courses different, but the interest and enthusiasm of students varied significantly. For this reason, only the thematic content of the comments was considered for the content analysis portion of the paper.

Appendix IV

Instructor & Course Comparison

The instructors of the courses, while not identical of course, appeared relatively well-matched for a comparison. Both Dr. Mitchell (a woman) and Dr. Martin (a man) are qualified, competent individuals in their respective fields (Dr. Mitchell received a Ph.D. in 2012, prior to all evaluations in the dataset, and Dr. Martin received a Ph.D. in 2014, prior to approximately

half of his evaluations). They are of similar age (Dr. Mitchell is 1 year older than Dr. Martin). While it would be nearly impossible to accurately measure differences in competency, attractiveness, or demeanor, the instructors both believed themselves to be rather well-liked and popular professors in their face-to-face courses, a sentiment echoed by the department's advisor, who is, as much as anyone can be, an expert on student preferences.

In Spring 2015, both Dr. Mitchell and Dr. Martin acted as instructor of record for several online introductory political science courses. Each course was identical in number/title, content, workload, and appearance. This particular introductory course is a required core course for all students seeking an undergraduate degree at a public university in this state. Enrollment in each section of the course was approximately 200-250.

All sections were open for registration simultaneously. Students had a tendency to enroll in the sections with the lowest number initially (merely because those sections appeared first in the registration list). This means that section 1 tended to fill up earlier than section 3 or 4. It may also be likely that students who enroll in courses early are systematically different than those who enroll later in the registration period; for example, they may be seniors, athletes, or simply motivated students. For this reason, we examined sections in the mid- to high- numerical order: sections 6, 7, 8, 9, and 10.

The courses each required students to complete nine participation quizzes (readings-based quizzes with two attempts having identical questions, worth a total of 30% of a student's final grade), eight discussion post assignments (discussion forum posts of approximately 350 words for each assignment, worth a total of 30% of a student's final grade), two short answer assignments (low-level analytical assignments approximately 900 words each, worth a total of 20% of a student's final grade), and two multiple choice exams (based on readings, worth a total

of 20% of a student's final grade). Non-graded components of the course included lessons (videos, timelines, or simulation activities) and readings (chapters in a textbook).

Written work was graded by graduate students in the political science or public administration program. All graders were given the same training and provided with the same rubric for grading assignments. Each of the five courses used in this analysis had a unique grader. Students were instructed to contact their grader with questions on written assignments. All sections of the course shared a course assistant. The course assistant was responsible for answering administrative or technical questions. Students were instructed to contact the course assistant first with general questions about the course. The course assistant was given a rigid set of course policies and adhered to them for all sections of the course, including the five sections involved in this research.

The instructor of each section sent an identical Welcome Message to students on the first day of classes with instructions on how to access the course, purchase materials, and get started. The instructors held weekly office hours for three hours per week, though Dr. Mitchell was typically in her office for more than the listed three hours. Both Dr. Mitchell and Dr. Martin frequently emailed with students and met with students. Both instructors most often responded to emails within 2-3 hours of receiving them, and as a rule, emails during the week received replies within 24 hours, and emails received on weekends received replies within 48 hours. Both instructors adhered to the same department-wide policies related to course issues such as late work, student emergencies, grading policies, and final grade rounding or curves.

Appendix V

Sample Emails

Student E-mail:

Professor Martin,

My name is Student Sample, and I am in your political science Intro online class. I made a mistake by uploading the discussion responses the way that I did. I bulls****ed the first responses in thinking that this assignment was something less than what it was. I feel bad with my initial response being less than honest. These open discussions with my classmates are obviously to create a platform of self reflection which I should take advantage of. Would you consider deleting the first two posts I commented on? I would really appreciate it.

Thank you for your time,
Student.

Professor Martin's Response:

Hi Student,

I can delete your responses, but it's a little tricky since there are so many posts in the forum. If you can let me know the name of the student who started the threads you posted in, I can go in and delete or edit the posts.

Student E-mail:

Hello Dr. Mitchell,

I am concerned about my grade. I have contacted the grader before for assistance, but the response I got was not helpful. My grades say that I am being counted off for not making word count, but I have made word count and included citations in my discussions and responses. What do you think my best plan of action would be? I can come into office hours or do anything that would help. I am striving for an A so I am willing to do whatever it takes. I received two 68s and on 97 on assignments, but I wrote all of them the exact same way. I am confused on why my writing is not sufficient for an A. Any advice you have, or time to work with me would be much appreciated. Thank you for your time.

Sincerely,

Student Sample

Professor Mitchell's Response:

Hello Student,

I appreciate that you're interested in improving your grade.

In your first discussion assignments, your initial posts were good, but your responses were not comprehensive and lacked citations. For your short answer, you needed to be sure to address the prompt fully and cite properly.

Your post and responses in Assignment 3 were much more complete and cited appropriately, which resulted in a much higher grade. Keep it up!

Student E-mail:

I received an error message when I sent this to you the first time, so I am not sure if the message made it through. I am sorry if it did; and that this might be a second copy, but I wanted to make sure that you received the work that I had gotten done. Thanks again for considering this, and I promise that this will not happen again.

Professor Mitchell's Response:

Unfortunately this document was not created prior to the due date and cannot be accepted for credit.

Appendix VI

Further Explanation of Methods in Ordinal Evaluation Comparison

In our comparison of ordinal evaluations, we intentionally chose to compare means (averages) rather than medians. Because the scale is relatively small (1-5), using the median would allow for very little variation across categories. In addition, comparing means allows the effects of outliers to be more apparent in the data.

Appendix VII

Unpaired T Test of SET by Question

Type of Question	
Instructor	5 questions, male instructor rated higher
	0 questions with insignificant difference
	0 questions, female instructor rated higher
Instructor/Course	5 questions, male instructor rated higher
	0 questions with insignificant difference
	0 questions, female instructor rated higher
Course	4 questions, male instructor rated higher
	3 questions with insignificant difference

	0 questions, female instructor rated higher
Technology	1 question, male instructor rated higher
	2 questions with insignificant difference
	0 questions, female instructor rated higher
Administrative	0 questions, male instructor rated higher
	3 questions with insignificant difference
	0 questions, female instructor rated higher
Total:	15 questions, male instructor rated higher
	8 questions with insignificant difference
	0 questions, female instructor rated higher