**Effects of Gender and STEM on Frequencies of Adjectives used in Online Reviews of Liberal Arts College Professors**

This study examines the influence of a professor’s gender (male or female) and their field (STEM or non-STEM) on the adjectives students used to describe them on RateMyProfessor. We scraped data from all the comments about professors belonging to the top twenty liberal arts colleges. Based on previous related literature and the adjectives most frequently used to describe these liberal arts professors, we created three word sets: “Smart, Brilliant, Knowledgeable,” “Hard, Difficult, Challenging,” and “Funny, Hilarious, Entertaining.” We then used three three-way Analysis of Variance model (ANOVA) to analyze the frequencies that different adjectives appeared in relation to three categorical explanatory variables: gender, department, and college. Our findings showed that male professors at top liberal arts colleges were more likely to be described as “Smart, Brilliant, Knowledgeable” than female professors. In addition, non-STEM professors were more often described as “Smart, Brilliant, Knowledgeable,” and less often characterized as “Hard, Difficult, Challenging” than their STEM counterparts.

**Background and Significance**

A number of studies explore gender-biases in higher education based upon professor course evaluations. A 1982 study found that students rated female professors highly only if they made themselves very available to provide help, while aloof male professors received equally high ratings (Koblitz, 2015). Further, female math professors received harsher evaluations than male math professors if students found their classes challenging (Koblitz, 2015). More recent literature reveals the continuation of gender-biased professor reviews. In a 2014 study, half the professors teaching the same online courses presented virtually as the opposite gender (disguising themselves with gendered names). Across the board, students rated perceived male professors more highly, even though instructors taught identical courses (Kamenetz, 2016). A search engine created in 2015 scraped 14 million student reviews on RateMyProfessor, revealing, for example, that male professors are far more often described as “brilliant,” and “funny,” whereas female professors appear “caring,” and “abrasive” (Leber, 2015). In 2016, economist Ann Boring found that student evaluations of teaching “are more sensitive to gender bias and grade expectations than they are to teaching effectiveness,” and that this “varies by discipline.”

While research on gender biases in higher education clearly exists, very little scholarship examines this form of persistent gender bias in liberal arts colleges. Our study aims to start filling this gap in the literature.

**Methods**

*Data Collection and Processing*:

We scraped all the RateMyProfessor reviews (N=48977) about every professor from the top twenty liberal arts colleges (US News 2018 Top 20).[[1]](#footnote-1) We then cleaned the data by deleting the empty and “No comments” columns. After cleaning our data, we ran an algorithm that counted the number of gender pronouns (he/him/his or she/her/hers) students used to describe each professor. By examining the prevailing gender pronouns in each professor’s reviews, we identified his or her gender. For the purpose of our study, we excluded all of the professors for whom we failed to determine gender (N=48885 after cleaning). Then, we collected the department information for each professor (supplied by RateMyProfessor) and categorized them into the broader fields of STEM and non-STEM.[[2]](#footnote-2)

*Variables:*

Our explanatory variables consisted of gender (male or female), field (STEM or non-STEM), and college. Based on previous research on gender-biased language in professor reviews (Sprague, 2005), and the frequency certain adjectives appeared in our data, we decided to create three sets of word groups:

* Group 1: Smart, Brilliant, Knowledgeable
* Group 2: Hard, Difficult, Challenging
* Group 3: Funny, Hilarious, Entertaining

We then calculated three response variables based upon the frequency of each word group. Specifically, we divided the number of times a word within a group appeared by the total number of comments each professor received. This calculation accounted for the fact that some professors simply receive more online reviews than others.

*Analytical Methods:*

Then we conducted a three-way ANOVA test for each of the three response variables, using our categorical variables (gender, field, and college). Note that residual plots indicated a need for a log transformation on all three response variables.[[3]](#footnote-3)

**Results**

The ANOVA tables for each response variable is provided in Appendix 2. Although we used “college” as one of our explanatory variables, we decided to focus our discussion on the significant interactions between our other two categorical variables, gender and field.[[4]](#footnote-4)

*Smart, Brilliant, Knowledgeable*

We found that gender field (p-value<0.001) and gender (p-value=0.004) and are both significant in influencing the frequency of this word set in students’ reviews of professors. From the main effects plot to the left, we can see that male professors are more likely to be described as “Smart, Brilliant, Knowledgeable.” Also, these adjectives appear far more frequently in reviews for non-STEM professors than STEM professors. The Gender\*Field interaction was not significant.

*Hard, Difficult, Challenging*



For this word set, we found that field (p-value<0.001) is significant, but gender (p-value=0.483) is not. From the interval plot to the left, we can tell that there are significant differences between STEM and non-STEM professor reviews. The STEM professors are far more frequently described as “Hard, Difficult, and Challenging.” The Gender\*Field interaction was not significant.

*Funny, Hilarious, Entertaining*

For this word set, gender is significant (p-value<0.001) but field is not (p-value=0.478).

This interaction plot shows that, for this word set, there appears a significant difference between genders (male/female). Male professors are more often described as “Funny, Hilarious, or Entertaining.” Additionally, there appears no significant difference based on field (STEM or non-STEM). The Gender\*Field interaction had a p-value = 0.09.

**Discussion**

Our results reveal gender to be a significant variable on which adjectives students used in their RateMyProfessor reviews. Male professors in liberal arts colleges are more often described as “Smart, Brilliant, Knowledgeable” and “Funny, Hilarious, Entertaining” than female professors (p-value<0.001 in both word sets). Additionally, non-STEM professors are generally described as more “Smart, Brilliant, Knowledgeable,” and less “Hard, Difficult, Challenging” than STEM professors (p-value<0.001 in both word sets).

Our analysis faced some limitations. One limitation with our dataset included the timeframe in which this data was generated. RateMyProfessor peaked in popularity with college students between 2003 and 2004. Therefore, many of the comments we scraped are from that period of time. Thus, the conclusions that we drew from this dataset may not completely reflect the environment in liberal arts colleges today. Another limitation includes the anonymity of the student reviewer, which made it impossible to investigate how the identity of the student influenced their comment.

Our results led us to consider not only the gender-biases discussed by previous research (Mickes & Laura, 2011), but also the different perceptions of professors in STEM and non-STEM fields. We think that students’ perception of STEM professors as less “Smart, Brilliant, Knowledgeable” than non-STEM professors between fields may not be due to the actual intelligence of professors. For example, the students who enroll in STEM (disproportionately male) may be less inclined to describe a professor as intelligent (Economics & Statistics Administration, 2017).

Future studies could make this research more comprehensive by increasing the word sets used. Additionally, further research could expand the scope by analyzing the differences between liberal arts colleges and universities. In addition, we noticed that female professors are more frequently reviewed as “Smart, Brilliant, Knowledgeable” than male professors at two women’s colleges, Wellesley College and Smith College (see interaction plot in appendix). This finding goes against the grain of our general results, providing a new direction for future research about differences between women’s and mixed-gender colleges.

**References**

Basow, S. A. (1994). Student ratings of professors are not gender blind. *AWM Newsletter*,

*24*(5), 1.

Bennett, S. K. (1982). Student perceptions of and expectations for male and female

instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, *74*(2), 170.

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not

measure teaching effectiveness. *ScienceOpen Research*.

Davis, K. (2015, August 20). The One Word Men Never See In Their Performance Reviews.

Retrieved from

http://www.fastcompany.com/3034895/the-one-word-men-never-see-in-their-performance-reviews.

Economics & Statistics Administration. (n.d.). Retrieved from

http://www.esa.doc.gov/reports/women-stem-2017-update

Kamenetz, A. (2014, September 26). Student Course Evaluations Get An 'F'. Retrieved from

http://www.npr.org/sections/ed/2014/09/26/345515451/student-course-evaluations-get-an-f

Kamenetz, A. (2016, January 25). Why Female Professors Get Lower Ratings. Retrieved from

http://www.npr.org/sections/ed/2016/01/25/463846130/why-women-professors-get-lower-rating.

Koblitz, N. (n.d.). Are Student Ratings Unfair to Women? by Neal Koblitz, University of

Washington. Retrieved from http://www.awm-math.org/newsletter/199009/koblitz.html

Leber, J. (2015, February 12). The Words Female Professors Hardly Ever Hear From Student

Reviews. Retrieved from

http://www.fastcompany.com/3042275/the-words-female-professors-hardly-ever-hear-from-student-reviews.

Mickes, L., Walker, D. E., Parris, J. L., Mankoff, R., & Christenfeld, N. J. (2012). Who’s funny:

Gender stereotypes, humor production, and memory bias. *Psychonomic Bulletin & Review*, *19*(1), 108-112.

Schmidt, B. (2015). Gendered language in teacher reviews. *Bookworm Project, feb*, *41.*

Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we

can't count can hurt us. *Sex Roles*, *53*(11-12), 779-793.

**Appendix 1**

**Footnote 1:**

U.S. News 2018 Best National Liberal Arts College:

<https://www.usnews.com/best-colleges/rankings/national-liberal-arts-colleges>

We excluded the United States Military Academy from our data set because we think it is different from other Liberal Arts Colleges in the list, and it does not have sufficient comments on RateMyProfessor.com. We only analyzed the reviews from 19 colleges.

**Footnote 2:**

STEM Designated Degree Program List: <https://www.ice.gov/sites/default/files/documents/Document/2016/stem-list.pdf>

The list is created based on Eligible CIP Codes for the STEM OPT Extension of United States Homeland Security.

['Psychology', 'Biology', 'Chemistry','Physics', 'Computer Science', 'Mathematics',

'Science', 'Neuropsychiatry', 'Physics & Astronomy', 'Statistics', 'Mathematics & Statistics', 'Physics Astronomy', 'Medicine', 'Engineering', 'Biological Sciences', 'Biochemistry', 'Astronomy', 'Botany', 'Cognitive Science', 'Neuroscience']

**Code**

Code used to scrape and process data:

<https://github.com/flyingelment/RateMyProfessor>

**Appendix 2**

**Footnote 3:**

Since the normal probability plot shows that our dataset is not normally distributed, we decided to transform our frequencies by completing the ln(100 \* n + 1) calculation on each of the professors. The way we calculated the frequencies caused the non-normal distribution of our data, since many comments lacked adjectives from the word sets, causing many zeros in our results. Even though the transformation did help somewhat, the graphs below show that our normality plots were still far from ideal. Since we have a large data set (N=48885), we believe that our test results are still valid.

**Data analysis:**

**Smart Brilliant Knowledgeable (1)**



Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

 College 18 16.00 0.8891 2.35 0.001

 Field 1 18.58 18.5792 49.18 0.000

 Gender 1 3.13 3.1275 8.28 0.004

 College\*Field 18 6.92 0.3846 1.02 0.435

 College\*Gender 18 12.70 0.7053 1.87 0.014

 Field\*Gender 1 0.02 0.0214 0.06 0.812

Error 5218 1971.14 0.3778

 Lack-of-Fit 18 1.94 0.1078 0.28 0.999

 Pure Error 5200 1969.20 0.3787

Total 5275 2045.41

In addition to the Field and Gender Main Effects, College and the College\*Gender Interaction were also significant. The above graph helps to visualize this interaction effect.

**Data analysis:**

**Smart Brilliant Knowledgeable (2)**



**Data analysis:**

**Hard difficult challenging**



Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

 College 18 115.39 6.4105 12.92 0.000

 Field 1 29.92 29.9186 60.31 0.000

 Gender 1 0.24 0.2446 0.49 0.483

 College\*Field 18 11.55 0.6419 1.29 0.180

 College\*Gender 18 7.12 0.3953 0.80 0.706

 Field\*Gender 1 0.49 0.4905 0.99 0.320

Error 5218 2588.52 0.4961

 Lack-of-Fit 18 9.80 0.5444 1.10 0.347

 Pure Error 5200 2578.72 0.4959

Total 5275 2798.85

**Data analysis:**

**Funny Hilarious Entertaining**



Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

 College 18 17.61 0.9781 2.68 0.000

 Field 1 0.18 0.1833 0.50 0.478

 Gender 1 52.24 52.2350 143.34 0.000

 College\*Field 18 7.74 0.4300 1.18 0.268

 College\*Gender 18 5.12 0.2846 0.78 0.725

 Field\*Gender 1 1.05 1.0455 2.87 0.090

Error 5218 1901.49 0.3644

 Lack-of-Fit 18 4.18 0.2320 0.64 0.874

 Pure Error 5200 1897.32 0.3649

Total 5275 2020.65

1. See Appendix 1 for ranking of top twenty liberal arts colleges. [↑](#footnote-ref-1)
2. See Appendix 1 for more detail. [↑](#footnote-ref-2)
3. See Appendix 2 for more information on our data transformation. [↑](#footnote-ref-3)
4. For more information on the significance of “college” as a variable see Appendix 2. [↑](#footnote-ref-4)