

Below is the manuscript that I submitted to PS: Political Science & Politics in September 2019, but with formatting changes such as spacing and converting endnotes to footnotes. This manuscript was rejected in January 2020 after peer review. I am not sure that the corrigendum for "Gender Bias in Student Evaluations" that was published in May 2020 is complete: for example, I think that the t-statistics in Table 4 of "Gender Bias in Student Evaluations" are incorrectly based on a calculation in which multiple ratings by the same student are treated as independent ratings. This manuscript contains more discussion of "Gender Bias in Student Evaluations" than is in the published corrigendum. Thanks to PS: Political Science & Politics for comments on my initial submission.

Comment on "Gender Bias in Student Evaluations"

L.J Zigerell
Associate Professor of Politics and Government
Illinois State University
Schroeder Hall 401
Normal IL 61790-4540
ljzigerell@ilstu.edu
@LJZigerell

Abstract

Mitchell and Martin (2018) reported evidence that student evaluations were biased against a female instructor. However, a reading of the article and its appendix and an analysis of the data indicate that there are important errors in the analysis and shortcomings in the research design that undercut inferences of gender bias in student evaluations of teaching based on the data reported on in the article.

Keywords: gender; gender bias; student evaluations; political science

Acknowledgements. The author thanks Dr. Kristina Mitchell for providing data and information for Mitchell and Martin (2018), thanks representatives from the Texas Tech University IRB and the Illinois State University IRB for respectively providing and forwarding the link to the Texas Tech student evaluations data, and thanks staff of the Texas Tech University Department of Political Science, representatives from Texas Tech University Human Resources, the Texas Tech University Office of General Counsel, and Kerri Ford of Texas Tech University Institutional Research and Information Management for information and/or assistance.

Comment on "Gender Bias in Student Evaluations"

Mitchell and Martin (2018) "Gender Bias in Student Evaluations" (hereafter MM) has been cited as evidence that student evaluations of teaching are biased against female instructors (e.g., Peterson et al. 2019, Sen 2018, and Sterett and Diascro 2019). MM has received a relatively large amount of attention, as the *PS: Political Science & Politics* article with the highest Altmetric score as of 25 July 2018 (Gronke et al. 2018). However, MM, the MM appendix, data on Texas Tech University website, and data provided to me by Dr. Mitchell (MM data) indicated that MM reported stronger evidence of gender bias than permitted by the MM data.

MM Content Analysis of Comments

MM Tables 1 and 2 reported a comparison of comments for MM co-author Kristina M. W. Mitchell to comments for MM co-author Jonathan Martin, respectively drawn from official university course evaluations and the Rate My Professors website. The table notes respectively indicate "N=68" and "N=54" but do not indicate whether these are total sample sizes or per-instructor sample sizes; the MM data indicate that these Ns are the total number of comments in each analysis, with respective counts of 45 comments for Mitchell and 23 comments for Martin and 45 comments for Mitchell and 9 comments for Martin.

Table 1 of the present manuscript reports results from MM Tables 1 and 2 and from my reanalysis based on my codings of the percentage of unique comments that met each theme.¹ My codings were necessitated because the data available to me did not indicate which comments MM coded to have met each theme and because percentages in MM Tables 1 and 2 could not be multiplied by the corresponding sample sizes to derive the number of unique comments that MM coded as meeting each theme.² I did not use a formula to determine inferential statistics for a test comparing the proportion of comments for the male instructor that met each theme to the proportion of comments for the female

¹ My coding followed as best I could the descriptions and examples in the MM appendix, except for the incompetency theme, which I limited to student comments about an instructor's subject matter knowledge, which seemed to be more in line with the percentages reported in MM Tables 1 and 2. Counts for "professor" included misspellings such as "proffessor", but did not include the one Rate My Professors comment referring to the female instructor as "professor" with sneer quotes; counts for "teacher" included misspellings such as "is a touch teach", but did not include comments in which it would be ungrammatical to substitute an alternate label such as "professor" (e.g., "is trying to teach"). The entertainment theme had close calls that could push the p-value under 0.05 or push the p-value higher than the p-value reported in the table; the coding for the reported analysis included "cool" as meeting the entertainment theme (for Martin's count) and as "entreating" being intended as "interesting" (for Mitchell's count).

² For example, MM Table 2 reports that 1.1% of comments for Mitchell referred to her intelligence or competency and that 20.9% of comments for Mitchell referred to her personality, which respectively correspond to about 0.5 of 45 comments and about 9.4 of 45 comments. The 6.6% of comments for Mitchell that referred to incompetency corresponds to about 3 of 45 comments for Mitchell, but, the potential for fractional coding means that this could have resulted from 3 comments counted as full comments or from a combination of more than 3 full or partial comments. Inferential statistics should consider 3 comments differently than, say, 2 full comments and 2 comments coded as half comments, thus necessitating my coding of the comments.

instructor that met each theme, because sample sizes were under 30 for the male instructor for each set of comments (Zou et al. 2003: 611). Instead, the p-value reported for my analysis indicates the result of a simulation in R (R Core Team 2018).

[Table 1 about here]

For example, my coding of the official university course evaluations comments indicated that 12 of 45 comments for Mitchell referred to her as a teacher (26.7%) and that 4 of 23 comments for Martin referred to him as a teacher (17.4%), a difference of 9.3% that is close to the 9.2% difference reported in MM Table 1. The 0.30 p-value for my analysis indicates that randomly drawing 45 comments for a female instructor from a pool that contained 16 "teacher" comments and a residual 52 non-"teacher" comments resulted in the female instructor receiving at least 12 "teacher" comments in 30% of ten million simulations. However, MM indicated a p-value under 0.05 for the "teacher" theme for official university course evaluations.

Table 1 of the present manuscript indicates that simulations for other comparisons also produced the inference that MM substantially overestimated the evidence of bias against the female instructor in the MM comments data.

MM discussed a "qualifications stereotype" that supported the expectation that "men are more likely to be referred to as 'professors' in their [student evaluations of teachers] and women are more likely to be referred to as 'teachers'" (p. 649). However, analyzing use of "teacher" in isolation might be misleading, given that "instructor" is plausibly similar in status to "teacher"; moreover, "Dr." can plausibly be considered to suggest qualifications as similarly high as "professor". My coding of the official university course evaluations indicated that none of the 23 comments for Martin referred to him as "Dr." but that 7 of the 45 comments for Mitchell referred to her as "Dr."; in contrast and counting misspellings, "Professor Martin" appeared in 3 of 23 comments and "Professor Mitchell" appeared in 1 of 45 comments.³ My coding of the official university course evaluations indicated that 8 comments for Martin referred to him as a Dr., Professor, or a professor, compared to 7 comments that referred to him as a teacher or an instructor, such that 53.3% of these "qualifications stereotype" comments for Martin were "higher qualifications" comments ("Dr.", "Professor", or "professor"); for Mitchell, the respective counts were 12 and 13, such that 52.0% of these "qualifications stereotype" comments for Mitchell were "lower qualifications" comments ("teacher" or "instructor"). Simulations indicated that this

³ The MM appendix and data indicated that these official university course evaluations comments were for different courses for Mitchell (POLS 3366, POLS 3361, and POLS 3312) and for Martin (POLS 3371 and POLS 3373). Texas Tech records indicated that Mitchell and Martin taught courses with these numbers in Spring 2015 or Fall 2015. Vitae online indicated that Mitchell and Martin respectively earned a Ph.D. in 2012 and 2014 and that Martin's teaching experiences began in 2013. My review of the Texas Tech University schedules for fall, summer, and spring semesters from Fall 2012 through Fall 2014 indicated that Martin had not taught POLS 3371 and POLS 3373 during those semesters. The December 2014 Texas Tech Commencement program listed Martin as earning a Ph.D., and 5 of the 9 Rate My Professors comments for Martin date to November 2014 or December 2014, so results are not reported for a simulation based on Rate My Professors comments that include "Dr.".

disadvantage for the female instructor or a larger disadvantage for the female instructor would happen about 3 in 10 times if these "qualifications" comments were randomly assigned; in particular, randomly drawing 45 comments for a female instructor from a pool that contained 20 "higher qualifications" comments and 48 residual comments and randomly drawing 45 comments for a female instructor from a pool that contained 20 "lower qualifications" comments and 48 residual comments (with remaining comments assigned to the male instructor) resulted in at least 52.0% of the qualifications comments for the female instructor being "lower qualifications" comments and at least 53.3% of the qualifications comments for the male instructor being "higher qualifications" comments in 31% of 10 million simulations.⁴

MM Analysis of Likert-Scale Student Ratings

Errors also appear in inferential statistics in the MM Table 4 analysis of student Likert-scale ratings for the 23 evaluation items for an online POLS 2302 American Public Policy course that Mitchell and Martin taught sections of in Spring 2015 at Texas Tech University. For the MM Table 4 analysis, these 23 evaluation items were grouped, such as 5 items grouped as comments on the instructor and 7 items grouped as comments on the course. This grouping explains the variation in MM Table 4 sample sizes: Mitchell and Martin respectively had Ns of 835 and 255 for the "instructor" analysis and Ns of 1,169 and 357 for the "course" analysis, which correspond to sample sizes of 167 students and 51 students when respectively dividing the pairs by 5 and by 7. However, inferential statistics will be incorrect if an analysis treats 167 students providing 7 ratings as if these were 1,169 independent ratings. My analysis indicated that MM Table 4 inferential statistics were calculated as if 7 ratings from one student were treated as 7 students each providing an independent rating.

Table 2 of the present manuscript reports individual-item results for student evaluations grouped across Mitchell's sections for which data are available compared to student evaluations grouped across Martin's sections; this individual-item analysis avoids the concern about multiple ratings from the same student being treated as independent ratings. This individual-item analysis provided little evidence of a difference in student ratings on some items, such as item 17 about the relevance and usefulness of the textbook and other purchased materials. However, the individual-item analysis did provide evidence at $p < 0.05$ that Mitchell and her sections for which data are available received lower ratings than did Martin and his sections, on the first and presumably most important instructor item ("Overall the instructor was effective") and on the first and presumably most important course item ("Overall, this course was a valuable learning experience"), at least when the ratings are grouped across sections in t-tests that do not assume equal variances by instructor. However, Table 2 indicates that p-values differed in linear regressions that clustered ratings by section of the course.

[Table 2 about here]

⁴ Qualifications comments could overlap, such as "Proffessor...is a phenomenal teacher".

Discussion

Endorsed by 17 professional organizations, the 2019 American Sociological Association Statement on Student Evaluations of Teaching claimed that "...students use stereotypically gendered language in how they evaluate their instructors", providing as evidence only Mitchell and Martin (2018). Parties attempting to make such an inference based on official university course evaluations comments and Rate My Professors comments reported on in Mitchell and Martin (2018) are cautioned about the overstated evidence discussed above and the fact that the data for male instructors were respectively only 23 comments for one male instructor and 9 comments for the same male instructor, both from non-experimental analyses.

The individual-item analysis of the Likert-scale student ratings data do provide evidence that, for many items, Mitchell and her sections of the course for which data are available received lower ratings than did Martin and his sections of the course. But parties attempting to make inferences about gender bias in student evaluations based on these data are cautioned about research design shortcomings that undercut the ability to assign these lower ratings to unfair gender bias among students. For example, MM provided no indication that students were randomly assigned to Mitchell or to Martin to help eliminate on-average differences between the students in Mitchell's sections and the students in Martin's sections. The MM appendix correctly indicates that student self-selection into sections can threaten the internal validity of the study: "It may also be likely that students who enroll in courses early are systematically different than those who enroll later in the registration period" (p. 4)⁵; remarkably, though, of the sections for which ratings data are available, Mitchell taught sections 6, 7, 8, and 9, and Martin taught sections 10 and 11, and it is unclear how, if at all, results were biased by the male instructor being assigned to higher section numbers than the female instructor was. Moreover, Mitchell taught four sections for which data are available and Martin taught two sections, but Texas Tech records indicated that Mitchell had more than three times as many students complete evaluations for her sections for which data are available and that, compared to Martin, Mitchell had a higher evaluation response rate across her sections for which data are available ($p < 0.01$); it is unclear how, if at all, these class size and response rate differences biased results.

⁵ Blåsjö (2018) expressed concerns about MM, including Mitchell and Martin serving in dual roles as the instructors who were evaluated by the students and as the researchers who evaluated bias against the instructors. Blåsjö (2018) also indicated that "the researchers threw away at least half of the actual data", in reference to the MM appendix indicating that MM reports Table 4 results only for section 6 through 10 of the course; however, it is not clear whether data for sections 1 through 5 of the course were collected: student evaluation data are available on the Texas Tech University website only for sections 6 through 11 of the course, and an open records request did not produce student evaluation data for sections 1 through 5. Related to the reference to sections 6 through 10, the MM appendix listing of evaluated sections as 6 through 10 appears to be an error and should have been listed as sections 6 through 11; a similar error regarding the number of sections for which data were analyzed appeared in MM: "We compared the ordinal evaluations of a man versus a woman in five sections of the online courses" (p. 650).

But even had students been randomly assigned to Mitchell or to Martin, Mitchell and Martin were not necessarily all else equal but for their gender in Spring 2015 at Texas Tech University. Texas Tech records indicate that, before the Spring 2015 semester in which the Likert scale analysis was conducted, Mitchell and Martin each had taught at least one Texas Tech political science course for at least three prior semesters. Therefore, during this prior experience, Mitchell and/or Martin could have developed reputations that influenced student evaluations for their Spring 2015 POLS 2302 sections, through direct prior experience with students who enrolled in their Spring 2015 POLS 2302 sections or indirectly through word-of-mouth or other reputation development mechanisms.⁶ Therefore, even if ratings in Mitchell's sections for which data are available differ from ratings in Martin's sections with a sufficiently low p-value, these differences cannot be assigned to only gender bias as if the analysis were a randomized experiment in which the female instructor was all else equal to the male instructor. Properly inferring gender bias in student evaluations of teaching from these data requires at least eliminating alternate explanations of observed differences being caused by student self-selection into sections and being caused by Mitchell and Martin having different reputations among the students before the semester started.

Conclusion

Nothing in this reanalysis indicates that student evaluations of teaching are not biased against female instructors, but the reanalysis does indicate that the MM data provide much less evidence for such gender bias than reported in Mitchell and Martin (2018).

References

- American Sociological Association. 2019. "Statement on Student Evaluations of Teaching". Retrieved from: https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_sept52019.pdf.
- Blåsjö, Viktor. 2018. "Notes on Mitchell & Martin (2018)". Intellectual Mathematics. March 20. <http://archive.is/7NGoT>.
- Gronke, Paul, Phillip Ardoin, and Celina Szymanski. 2018. "Editor's Report for *PS: Political Science & Politics*". *PS: Political Science & Politics* 51(4): 903-906.
- Mitchell, Kristina M. W., and Jonathan Martin. 2018. "Gender Bias in Student Evaluations". *PS: Political Science & Politics* 51(3): 1-5.
- Peterson, David A. M., Lori A. Biederman, David Andersen, Tessa M. Ditonto, and Kevin Roe. 2019. "Mitigating Gender Bias in Student Evaluations of Teaching". *PLOS ONE* 14(5): e0216241.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

⁶ As Mitchell and Martin (2018) noted: "...consistently poor evaluations in this [Rate My Professors] and similar public forums can have career implications for academics due to lower enrollments in courses or an unfavorable reputation on campus" (p. 649).

- Sen, Maya. 2018. "Response to Dion, Sumner, and Mitchell". *Political Analysis* 26(3): 335-337.
- Sterett, Susan M., and Jennifer A. Diascro. 2019. "Introduction: Reflecting on the Profession". *PS: Political Science & Politics* 52(1): 31-34.
- Zou, Kelly H., Julia R. Fielding, Stuart G. Silverman, and Clare MC Tempany. 2003. "Hypothesis Testing I: Proportions". *Radiology* 226(3): 609-613.

Table 1						
Reanalysis of Student Comments for Selected Themes from Mitchell and Martin (2018) Tables 1 and 2						
	MM Analysis		Reanalysis			
	Percentage point male/female difference	p-value	Male teacher count	Female teacher count	Percentage point male/female difference	p-value
Official University Course Evaluations						
Personality	-11.2	p<0.01	1 of 23	8 of 45	-13.4	p=0.12
Appearance	--	--	--	--	--	--
Entertainment	-17.0	p<0.01	3 of 23	15 of 45	-20.3	p=0.06
Intelligence/Competency	--	--	--	--	--	--
Incompetency	--	--	--	--	--	--
Referred to as "professor"	17.1	p<0.01	8 of 23	7 of 45	19.2	p=0.07
Referred to as "teacher"	-9.2	p<0.05	4 of 23	12 of 45	-9.3	p=0.30
Rate My Professors Comments						
Personality	-9.9	p<0.05	1 of 9	11 of 45	-13.3	p=0.35
Appearance	-10.6	p<0.05	0 of 9	5 of 45	-11.1	p=0.39
Entertainment	--	--	--	--	--	--
Intelligence/Competency	--	--	--	--	--	--
Incompetency	-6.6	p<0.10	0 of 9	3 of 45	-6.6	p=0.57
Referred to as "professor"	--	--	--	--	--	--
Referred to as "teacher"	-5.5	p<0.05	0 of 9	4 of 45	-8.9	p=0.47

Note: Data for the left two numeric columns are drawn from Mitchell and Martin (2018). Data from the right four numeric columns are drawn from the author's reanalysis of the MM data. Dashed lines indicate comparisons for which MM did not report a statistical significance asterisk. P-values in the rightmost column are based on simulations that estimated how often the female instructor would have received at least as many comments as were observed on her personality, on her appearance, on the entertainment aspect of her or her course, on her incompetency, and that referred to her as a teacher, and at least as few comments as were observed that referred to her as professor or a professor, given random assignment of the number of observed comments for each theme based on the number of observed comments for the female instructor and the number of observed comments for the male instructor.

#	Item text	Mean for Mitchell (std dev) N=167	Mean for Martin (std dev) N=51	Two-tailed p-value	
				Not clustered by section	Clustered by section
1	Overall the instructor(s) was (were) effective.	3.24 (1.21)	3.69 (1.09)	0.01	0.09
2	The instructor(s) was (were) available for consultation in a timely manner.	3.57 (0.98)	3.94 (0.90)	0.01	0.04
3	The instructor(s) stimulated student learning.	3.17 (1.20)	3.63 (1.22)	0.02	0.05
4	The instructor(s) treated me fairly.	3.50 (1.10)	3.84 (0.97)	0.03	0.12
5	The instructor(s) treated me with respect.	3.54 (1.07)	3.98 (0.93)	0.01	0.06
6	The instructor(s) welcomed and encouraged questions and comments.	3.35 (1.17)	3.78 (1.03)	0.01	0.01
7	The instructor(s) presented the information clearly.	3.49 (1.19)	3.92 (0.96)	0.01	0.13
8	The instructor(s) emphasized the major points and concepts.	3.47 (1.15)	3.76 (1.03)	0.08	0.16
9	The instructor(s) went beyond presenting the information in the text.	3.07 (1.28)	3.53 (1.15)	0.02	0.04
10	The instructor(s) demonstrated knowledge of the subject.	3.39 (1.16)	3.69 (1.03)	0.08	0.05
11	Overall this course was a valuable learning experience.	3.16 (1.33)	3.63 (1.23)	0.02	0.03
12	The assignments were relevant and useful.	3.37 (1.20)	3.65 (1.15)	0.14	0.10
13	Course materials were relevant and useful.	3.50 (1.17)	3.65 (1.20)	0.43	0.27
14	Expectations were clearly stated either verbally or in the syllabus.	3.67 (1.25)	4.00 (0.92)	0.04	0.17
15	The testing and evaluation procedures were fair.	3.40 (1.28)	3.74 (1.15)	0.07	0.11
16	The workload was appropriate for the hours of credit.	3.63 (1.15)	3.71 (1.04)	0.68	0.48
17	The textbook or other purchased materials were relevant and useful.	3.68 (1.17)	3.63 (1.13)	0.79	0.72
18	The technology for delivering the instruction was appropriate for the course.	3.74 (1.06)	3.82 (0.99)	0.62	0.44

19	Information was available to help me solve technical problems before they caused delays in my course	3.50 (1.16)	3.84 (0.97)	0.04	0.02
20	Documentation for accessing electronically-delivered class lectures and/or material was effective.	3.68 (1.04)	3.82 (0.95)	0.35	0.23
21	Registration procedures for the course ran smoothly.	4.13 (0.84)	4.10 (0.81)	0.83	0.84
22	Advising was available and adequate.	3.75 (1.07)	3.84 (0.97)	0.55	0.42
23	Course materials were readily accessible.	4.03 (0.98)	3.94 (0.95)	0.56	0.58

Online Appendix for Comment on "Gender Bias in Student Evaluations"

R code for the analysis...

```
#####  
# MM Table 1 #  
#####
```

```
# University evaluations "personality" comments  
# Simulation of how often, the female instructor would receive 8 or more "personality"  
# comments, if 9 "personality" comments were randomly assigned to a female instructor  
# who received 45 comments and a male instructor who received 23 comments.
```

```
n.mitchell      <- 45  
n.martin        <- 23  
n.items         <- 9  
counter         <- 0  
n.simulations   <- 10000000  
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {  
  mitchell <- sample(comments, n.mitchell, replace=FALSE)  
  if (sum(mitchell)>=8) {  
    counter <- counter + 1  
  }  
}  
counter/n.simulations
```

```
# University evaluations "entertainment" comments  
# Simulation of how often, the female instructor would receive 15 or more "entertainment"  
# comments, if 18 "entertainment" comments were randomly assigned to a female instructor  
# who received 45 comments and a male instructor who received 23 comments.
```

```
n.mitchell      <- 45  
n.martin        <- 23  
n.items         <- 18  
counter         <- 0  
n.simulations   <- 10000000  
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {  
  mitchell <- sample(comments, n.mitchell, replace=FALSE)  
  if (sum(mitchell)>=15) {  
    counter <- counter + 1  
  }  
}  
counter/n.simulations
```

```
# University evaluations "professor" comments
# Simulation of how often, the female instructor would receive 7 or fewer "professor"
# comments, if 15 "professor" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 23 comments.
```

```
n.mitchell      <- 45
n.martin        <- 23
n.items         <- 15
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)<=7) {
    counter <- counter + 1
  }
}
counter/n.simulations
```

```
# University evaluations "teacher" comments
# Simulation of how often, the female instructor would receive 12 or more "teacher"
# comments, if 16 "teacher" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 23 comments.
```

```
n.mitchell      <- 45
n.martin        <- 23
n.items         <- 16
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)>=12) {
    counter <- counter + 1
  }
}
counter/n.simulations
```

```

#####
# MM Table 2 #
#####

# Rate My Professors "personality" comments
# Simulation of how often, the female instructor would receive 11 or more "personality"
# comments, if 12 "personality" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 9 comments.

n.mitchell      <- 45
n.martin        <- 9
n.items         <- 12
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))

for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)>=11) {
    counter <- counter + 1
  }
}
counter/n.simulations

# Rate My Professors "appearance" comments
# Simulation of how often, the female instructor would receive all 5 "appearance"
# comments, if 5 "appearance" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 9 comments.

n.mitchell      <- 45
n.martin        <- 9
n.items         <- 5
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))

for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)==5) {
    counter <- counter + 1
  }
}
counter/n.simulations

```

```
# Rate My Professors "incompetency" comments
# Simulation of how often, the female instructor would receive all 3 "incompetency"
# comments, if 3 "incompetency" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 9 comments.
```

```
n.mitchell      <- 45
n.martin        <- 9
n.items         <- 3
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)==3) {
    counter <- counter + 1
  }
}
counter/n.simulations
```

```
# Rate My Professors "teacher" comments
# Simulation of how often, the female instructor would receive all 4 "teacher"
# comments, if 4 "teacher" comments were randomly assigned to a female instructor
# who received 45 comments and a male instructor who received 9 comments.
```

```
n.mitchell      <- 45
n.martin        <- 9
n.items         <- 4
counter         <- 0
n.simulations   <- 10000000
comments        <- c(rep_len(1, n.items), rep_len(0, n.mitchell+n.martin-n.items))
```

```
for (i in 1:n.simulations) {
  mitchell <- sample(comments, n.mitchell, replace=FALSE)
  if (sum(mitchell)==4) {
    counter <- counter + 1
  }
}
counter/n.simulations
```

```

#####
# Not from MM #
#####

# University evaluations "professor" and "Dr." comments
# compared to "teacher" and "instructor" comments

# Simulation of how often the female instructor would receive a percentage of
# "qualifications stereotype comments" that was at least 13/25ths "lower qualifications"
# comments and the male instructor would receive a percentage of "qualifications
# stereotype comments" that was at least 8/15ths "higher qualifications" comments, if
# 20 "lower qualifications" comments and 20 "higher qualifications" comments were
# randomly assigned to a female instructor who received 45 comments and a male instructor
# who received 23 comments.

n.mitchell      <- 45
n.martin        <- 23
n.ti            <- 20
n.pd           <- 20
n.simulations   <- 10000000
count          <- 0

comments.ti <- c(rep_len(1, n.ti), rep_len(0, n.mitchell+n.martin-n.ti))
comments.pd <- c(rep_len(1, n.pd), rep_len(0, n.mitchell+n.martin-n.pd))

for (i in 1:n.simulations){

  mitchell.ti <- sum(sample(comments.ti, n.mitchell, replace=FALSE))
  mitchell.pd <- sum(sample(comments.pd, n.mitchell, replace=FALSE))

  martin.ti   <- n.ti - mitchell.ti
  martin.pd   <- n.pd - mitchell.pd

  if (
    (mitchell.ti/(mitchell.ti+mitchell.pd) >= 13/25) &
    (martin.pd/(martin.ti+martin.pd) >= 8/15)
  )
  {
    count <- count + 1
  }
}

count/n.simulations

```

```
#####  
# MM Table 4 #  
#####
```

```
# Code to create the Table 4 database:  
# i01w06 = item 1 for the woman instructor for course section 6
```

```
i01w06 <- c(rep_len(5,15), rep_len(4,14), rep_len(3,19), rep_len(2,8), rep_len(1,2))  
i02w06 <- c(rep_len(5,16), rep_len(4,21), rep_len(3,18), rep_len(2,2), rep_len(1,1))  
i03w06 <- c(rep_len(5,13), rep_len(4,17), rep_len(3,15), rep_len(2,10), rep_len(1,3))  
i04w06 <- c(rep_len(5,13), rep_len(4,24), rep_len(3,18), rep_len(2,2), rep_len(1,1))  
i05w06 <- c(rep_len(5,14), rep_len(4,19), rep_len(3,20), rep_len(2,2), rep_len(1,3))  
i06w06 <- c(rep_len(5,15), rep_len(4,10), rep_len(3,25), rep_len(2,5), rep_len(1,3))  
i07w06 <- c(rep_len(5,15), rep_len(4,27), rep_len(3,12), rep_len(2,3), rep_len(1,1))  
i08w06 <- c(rep_len(5,15), rep_len(4,25), rep_len(3,13), rep_len(2,4), rep_len(1,1))  
i09w06 <- c(rep_len(5,12), rep_len(4,14), rep_len(3,14), rep_len(2,14), rep_len(1,4))  
i10w06 <- c(rep_len(5,14), rep_len(4,16), rep_len(3,19), rep_len(2,6), rep_len(1,3))  
i11w06 <- c(rep_len(5,13), rep_len(4,19), rep_len(3,13), rep_len(2,8), rep_len(1,5))  
i12w06 <- c(rep_len(5,14), rep_len(4,21), rep_len(3,14), rep_len(2,5), rep_len(1,4))  
i13w06 <- c(rep_len(5,11), rep_len(4,26), rep_len(3,13), rep_len(2,5), rep_len(1,3))  
i14w06 <- c(rep_len(5,19), rep_len(4,29), rep_len(3,5), rep_len(2,3), rep_len(1,2))  
i15w06 <- c(rep_len(5,16), rep_len(4,25), rep_len(3,10), rep_len(2,3), rep_len(1,4))  
i16w06 <- c(rep_len(5,17), rep_len(4,21), rep_len(3,14), rep_len(2,2), rep_len(1,4))  
i17w06 <- c(rep_len(5,17), rep_len(4,24), rep_len(3,12), rep_len(2,3), rep_len(1,2))  
i18w06 <- c(rep_len(5,17), rep_len(4,26), rep_len(3,10), rep_len(2,3), rep_len(1,2))  
i19w06 <- c(rep_len(5,15), rep_len(4,17), rep_len(3,20), rep_len(2,2), rep_len(1,4))  
i20w06 <- c(rep_len(5,15), rep_len(4,24), rep_len(3,15), rep_len(2,2), rep_len(1,2))  
i21w06 <- c(rep_len(5,20), rep_len(4,34), rep_len(3,2), rep_len(2,2), rep_len(1,0))  
i22w06 <- c(rep_len(5,18), rep_len(4,23), rep_len(3,14), rep_len(2,2), rep_len(1,1))  
i23w06 <- c(rep_len(5,23), rep_len(4,30), rep_len(3,4), rep_len(2,1), rep_len(1,0))  
i01w07 <- c(rep_len(5,7), rep_len(4,14), rep_len(3,9), rep_len(2,4), rep_len(1,6))  
i02w07 <- c(rep_len(5,6), rep_len(4,15), rep_len(3,17), rep_len(2,1), rep_len(1,1))  
i03w07 <- c(rep_len(5,5), rep_len(4,10), rep_len(3,13), rep_len(2,7), rep_len(1,5))  
i04w07 <- c(rep_len(5,7), rep_len(4,14), rep_len(3,14), rep_len(2,2), rep_len(1,3))  
i05w07 <- c(rep_len(5,7), rep_len(4,15), rep_len(3,16), rep_len(2,0), rep_len(1,2))  
i06w07 <- c(rep_len(5,5), rep_len(4,17), rep_len(3,10), rep_len(2,5), rep_len(1,3))  
i07w07 <- c(rep_len(5,8), rep_len(4,16), rep_len(3,9), rep_len(2,3), rep_len(1,4))  
i08w07 <- c(rep_len(5,8), rep_len(4,15), rep_len(3,8), rep_len(2,5), rep_len(1,4))  
i09w07 <- c(rep_len(5,4), rep_len(4,12), rep_len(3,10), rep_len(2,9), rep_len(1,5))  
i10w07 <- c(rep_len(5,7), rep_len(4,14), rep_len(3,14), rep_len(2,3), rep_len(1,2))  
i11w07 <- c(rep_len(5,4), rep_len(4,18), rep_len(3,7), rep_len(2,4), rep_len(1,7))  
i12w07 <- c(rep_len(5,6), rep_len(4,18), rep_len(3,8), rep_len(2,3), rep_len(1,5))  
i13w07 <- c(rep_len(5,6), rep_len(4,21), rep_len(3,8), rep_len(2,3), rep_len(1,2))  
i14w07 <- c(rep_len(5,12), rep_len(4,17), rep_len(3,6), rep_len(2,1), rep_len(1,4))  
i15w07 <- c(rep_len(5,7), rep_len(4,13), rep_len(3,11), rep_len(2,4), rep_len(1,5))  
i16w07 <- c(rep_len(5,9), rep_len(4,18), rep_len(3,9), rep_len(2,2), rep_len(1,2))  
i17w07 <- c(rep_len(5,10), rep_len(4,19), rep_len(3,5), rep_len(2,3), rep_len(1,3))  
i18w07 <- c(rep_len(5,7), rep_len(4,25), rep_len(3,4), rep_len(2,1), rep_len(1,3))  
i19w07 <- c(rep_len(5,10), rep_len(4,16), rep_len(3,8), rep_len(2,4), rep_len(1,2))  
i20w07 <- c(rep_len(5,8), rep_len(4,17), rep_len(3,13), rep_len(2,1), rep_len(1,1))  
i21w07 <- c(rep_len(5,16), rep_len(4,21), rep_len(3,2), rep_len(2,1), rep_len(1,0))  
i22w07 <- c(rep_len(5,13), rep_len(4,12), rep_len(3,11), rep_len(2,3), rep_len(1,1))  
i23w07 <- c(rep_len(5,17), rep_len(4,16), rep_len(3,5), rep_len(2,2), rep_len(1,0))  
i01w08 <- c(rep_len(5,5), rep_len(4,7), rep_len(3,18), rep_len(2,7), rep_len(1,5))  
i02w08 <- c(rep_len(5,7), rep_len(4,9), rep_len(3,22), rep_len(2,1), rep_len(1,3))  
i03w08 <- c(rep_len(5,5), rep_len(4,8), rep_len(3,17), rep_len(2,7), rep_len(1,5))
```

```

i04w08 <- c(rep_len(5,6), rep_len(4,8), rep_len(3,17), rep_len(2,5), rep_len(1,6))
i05w08 <- c(rep_len(5,7), rep_len(4,11), rep_len(3,17), rep_len(2,3), rep_len(1,4))
i06w08 <- c(rep_len(5,7), rep_len(4,10), rep_len(3,14), rep_len(2,5), rep_len(1,6))
i07w08 <- c(rep_len(5,6), rep_len(4,10), rep_len(3,12), rep_len(2,9), rep_len(1,5))
i08w08 <- c(rep_len(5,6), rep_len(4,11), rep_len(3,13), rep_len(2,8), rep_len(1,4))
i09w08 <- c(rep_len(5,5), rep_len(4,8), rep_len(3,11), rep_len(2,8), rep_len(1,10))
i10w08 <- c(rep_len(5,6), rep_len(4,10), rep_len(3,16), rep_len(2,2), rep_len(1,8))
i11w08 <- c(rep_len(5,7), rep_len(4,8), rep_len(3,8), rep_len(2,10), rep_len(1,9))
i12w08 <- c(rep_len(5,5), rep_len(4,11), rep_len(3,12), rep_len(2,10), rep_len(1,4))
i13w08 <- c(rep_len(5,8), rep_len(4,11), rep_len(3,9), rep_len(2,8), rep_len(1,6))
i14w08 <- c(rep_len(5,12), rep_len(4,5), rep_len(3,11), rep_len(2,6), rep_len(1,8))
i15w08 <- c(rep_len(5,9), rep_len(4,8), rep_len(3,10), rep_len(2,10), rep_len(1,5))
i16w08 <- c(rep_len(5,10), rep_len(4,14), rep_len(3,11), rep_len(2,4), rep_len(1,3))
i17w08 <- c(rep_len(5,11), rep_len(4,14), rep_len(3,11), rep_len(2,3), rep_len(1,3))
i18w08 <- c(rep_len(5,10), rep_len(4,13), rep_len(3,14), rep_len(2,3), rep_len(1,2))
i19w08 <- c(rep_len(5,8), rep_len(4,8), rep_len(3,17), rep_len(2,6), rep_len(1,3))
i20w08 <- c(rep_len(5,11), rep_len(4,9), rep_len(3,16), rep_len(2,4), rep_len(1,2))
i21w08 <- c(rep_len(5,14), rep_len(4,20), rep_len(3,4), rep_len(2,3), rep_len(1,1))
i22w08 <- c(rep_len(5,12), rep_len(4,10), rep_len(3,14), rep_len(2,3), rep_len(1,3))
i23w08 <- c(rep_len(5,13), rep_len(4,15), rep_len(3,8), rep_len(2,4), rep_len(1,2))
i01w09 <- c(rep_len(5,3), rep_len(4,3), rep_len(3,13), rep_len(2,3), rep_len(1,5))
i02w09 <- c(rep_len(5,2), rep_len(4,9), rep_len(3,12), rep_len(2,1), rep_len(1,3))
i03w09 <- c(rep_len(5,3), rep_len(4,5), rep_len(3,10), rep_len(2,4), rep_len(1,5))
i04w09 <- c(rep_len(5,4), rep_len(4,13), rep_len(3,6), rep_len(2,1), rep_len(1,3))
i05w09 <- c(rep_len(5,4), rep_len(4,12), rep_len(3,7), rep_len(2,1), rep_len(1,3))
i06w09 <- c(rep_len(5,4), rep_len(4,8), rep_len(3,9), rep_len(2,3), rep_len(1,3))
i07w09 <- c(rep_len(5,5), rep_len(4,8), rep_len(3,7), rep_len(2,2), rep_len(1,5))
i08w09 <- c(rep_len(5,3), rep_len(4,8), rep_len(3,9), rep_len(2,4), rep_len(1,3))
i09w09 <- c(rep_len(5,4), rep_len(4,8), rep_len(3,8), rep_len(2,2), rep_len(1,5))
i10w09 <- c(rep_len(5,4), rep_len(4,8), rep_len(3,10), rep_len(2,2), rep_len(1,3))
i11w09 <- c(rep_len(5,3), rep_len(4,8), rep_len(3,7), rep_len(2,2), rep_len(1,7))
i12w09 <- c(rep_len(5,4), rep_len(4,9), rep_len(3,7), rep_len(2,3), rep_len(1,4))
i13w09 <- c(rep_len(5,5), rep_len(4,12), rep_len(3,5), rep_len(2,1), rep_len(1,4))
i14w09 <- c(rep_len(5,5), rep_len(4,11), rep_len(3,6), rep_len(2,2), rep_len(1,3))
i15w09 <- c(rep_len(5,5), rep_len(4,7), rep_len(3,4), rep_len(2,7), rep_len(1,4))
i16w09 <- c(rep_len(5,3), rep_len(4,12), rep_len(3,6), rep_len(2,1), rep_len(1,5))
i17w09 <- c(rep_len(5,4), rep_len(4,11), rep_len(3,4), rep_len(2,2), rep_len(1,6))
i18w09 <- c(rep_len(5,4), rep_len(4,13), rep_len(3,5), rep_len(2,2), rep_len(1,3))
i19w09 <- c(rep_len(5,4), rep_len(4,10), rep_len(3,6), rep_len(2,3), rep_len(1,4))
i20w09 <- c(rep_len(5,4), rep_len(4,12), rep_len(3,6), rep_len(2,2), rep_len(1,3))
i21w09 <- c(rep_len(5,6), rep_len(4,13), rep_len(3,5), rep_len(2,2), rep_len(1,1))
i22w09 <- c(rep_len(5,4), rep_len(4,10), rep_len(3,10), rep_len(2,0), rep_len(1,3))
i23w09 <- c(rep_len(5,6), rep_len(4,12), rep_len(3,4), rep_len(2,2), rep_len(1,3))
i01m10 <- c(rep_len(5,6), rep_len(4,12), rep_len(3,10), rep_len(2,1), rep_len(1,3))
i02m10 <- c(rep_len(5,7), rep_len(4,17), rep_len(3,7), rep_len(2,0), rep_len(1,1))
i03m10 <- c(rep_len(5,5), rep_len(4,15), rep_len(3,6), rep_len(2,3), rep_len(1,3))
i04m10 <- c(rep_len(5,6), rep_len(4,16), rep_len(3,8), rep_len(2,0), rep_len(1,2))
i05m10 <- c(rep_len(5,6), rep_len(4,18), rep_len(3,6), rep_len(2,0), rep_len(1,2))
i06m10 <- c(rep_len(5,7), rep_len(4,13), rep_len(3,9), rep_len(2,2), rep_len(1,1))
i07m10 <- c(rep_len(5,8), rep_len(4,13), rep_len(3,8), rep_len(2,2), rep_len(1,1))
i08m10 <- c(rep_len(5,6), rep_len(4,15), rep_len(3,8), rep_len(2,2), rep_len(1,1))
i09m10 <- c(rep_len(5,5), rep_len(4,10), rep_len(3,13), rep_len(2,1), rep_len(1,3))
i10m10 <- c(rep_len(5,5), rep_len(4,16), rep_len(3,7), rep_len(2,3), rep_len(1,1))
i11m10 <- c(rep_len(5,7), rep_len(4,13), rep_len(3,8), rep_len(2,1), rep_len(1,3))
i12m10 <- c(rep_len(5,7), rep_len(4,13), rep_len(3,9), rep_len(2,1), rep_len(1,2))
i13m10 <- c(rep_len(5,7), rep_len(4,14), rep_len(3,8), rep_len(2,0), rep_len(1,3))
i14m10 <- c(rep_len(5,9), rep_len(4,17), rep_len(3,5), rep_len(2,0), rep_len(1,1))

```

```

i15m10 <- c(rep_len(5,7), rep_len(4,15), rep_len(3,6), rep_len(2,2), rep_len(1,2))
i16m10 <- c(rep_len(5,6), rep_len(4,16), rep_len(3,5), rep_len(2,4), rep_len(1,1))
i17m10 <- c(rep_len(5,7), rep_len(4,13), rep_len(3,8), rep_len(2,3), rep_len(1,1))
i18m10 <- c(rep_len(5,6), rep_len(4,16), rep_len(3,8), rep_len(2,1), rep_len(1,1))
i19m10 <- c(rep_len(5,6), rep_len(4,17), rep_len(3,8), rep_len(2,0), rep_len(1,1))
i20m10 <- c(rep_len(5,5), rep_len(4,18), rep_len(3,6), rep_len(2,2), rep_len(1,1))
i21m10 <- c(rep_len(5,9), rep_len(4,20), rep_len(3,3), rep_len(2,0), rep_len(1,0))
i22m10 <- c(rep_len(5,7), rep_len(4,17), rep_len(3,6), rep_len(2,1), rep_len(1,1))
i23m10 <- c(rep_len(5,8), rep_len(4,18), rep_len(3,3), rep_len(2,2), rep_len(1,1))
i01m11 <- c(rep_len(5,7), rep_len(4,5), rep_len(3,6), rep_len(2,1), rep_len(1,0))
i02m11 <- c(rep_len(5,8), rep_len(4,4), rep_len(3,6), rep_len(2,1), rep_len(1,0))
i03m11 <- c(rep_len(5,9), rep_len(4,2), rep_len(3,5), rep_len(2,2), rep_len(1,1))
i04m11 <- c(rep_len(5,7), rep_len(4,6), rep_len(3,5), rep_len(2,1), rep_len(1,0))
i05m11 <- c(rep_len(5,9), rep_len(4,6), rep_len(3,4), rep_len(2,0), rep_len(1,0))
i06m11 <- c(rep_len(5,8), rep_len(4,3), rep_len(3,6), rep_len(2,2), rep_len(1,0))
i07m11 <- c(rep_len(5,8), rep_len(4,6), rep_len(3,5), rep_len(2,0), rep_len(1,0))
i08m11 <- c(rep_len(5,8), rep_len(4,2), rep_len(3,8), rep_len(2,0), rep_len(1,1))
i09m11 <- c(rep_len(5,7), rep_len(4,4), rep_len(3,5), rep_len(2,2), rep_len(1,1))
i10m11 <- c(rep_len(5,6), rep_len(4,5), rep_len(3,6), rep_len(2,1), rep_len(1,1))
i11m11 <- c(rep_len(5,7), rep_len(4,4), rep_len(3,4), rep_len(2,2), rep_len(1,2))
i12m11 <- c(rep_len(5,6), rep_len(4,4), rep_len(3,6), rep_len(2,1), rep_len(1,2))
i13m11 <- c(rep_len(5,7), rep_len(4,2), rep_len(3,7), rep_len(2,1), rep_len(1,2))
i14m11 <- c(rep_len(5,6), rep_len(4,8), rep_len(3,4), rep_len(2,0), rep_len(1,1))
i15m11 <- c(rep_len(5,7), rep_len(4,5), rep_len(3,5), rep_len(2,0), rep_len(1,2))
i16m11 <- c(rep_len(5,6), rep_len(4,4), rep_len(3,8), rep_len(2,0), rep_len(1,1))
i17m11 <- c(rep_len(5,6), rep_len(4,3), rep_len(3,7), rep_len(2,1), rep_len(1,2))
i18m11 <- c(rep_len(5,7), rep_len(4,6), rep_len(3,4), rep_len(2,1), rep_len(1,1))
i19m11 <- c(rep_len(5,7), rep_len(4,5), rep_len(3,5), rep_len(2,1), rep_len(1,1))
i20m11 <- c(rep_len(5,6), rep_len(4,8), rep_len(3,4), rep_len(2,0), rep_len(1,1))
i21m11 <- c(rep_len(5,7), rep_len(4,6), rep_len(3,5), rep_len(2,0), rep_len(1,1))
i22m11 <- c(rep_len(5,6), rep_len(4,5), rep_len(3,7), rep_len(2,0), rep_len(1,1))
i23m11 <- c(rep_len(5,7), rep_len(4,5), rep_len(3,6), rep_len(2,1), rep_len(1,0))

```

```
# Code to check the numbers above:
```

```

rev(table(i01w06))
rev(table(i02w06))
rev(table(i03w06))
rev(table(i04w06))
rev(table(i05w06))
rev(table(i06w06))
rev(table(i07w06))
rev(table(i08w06))
rev(table(i09w06))
rev(table(i10w06))
rev(table(i11w06))
rev(table(i12w06))
rev(table(i13w06))
rev(table(i14w06))
rev(table(i15w06))
rev(table(i16w06))
rev(table(i17w06))
rev(table(i18w06))
rev(table(i19w06))
rev(table(i20w06))
rev(table(i21w06))
rev(table(i22w06))

```

rev(table(i23w06))
rev(table(i01w07))
rev(table(i02w07))
rev(table(i03w07))
rev(table(i04w07))
rev(table(i05w07))
rev(table(i06w07))
rev(table(i07w07))
rev(table(i08w07))
rev(table(i09w07))
rev(table(i10w07))
rev(table(i11w07))
rev(table(i12w07))
rev(table(i13w07))
rev(table(i14w07))
rev(table(i15w07))
rev(table(i16w07))
rev(table(i17w07))
rev(table(i18w07))
rev(table(i19w07))
rev(table(i20w07))
rev(table(i21w07))
rev(table(i22w07))
rev(table(i23w07))
rev(table(i01w08))
rev(table(i02w08))
rev(table(i03w08))
rev(table(i04w08))
rev(table(i05w08))
rev(table(i06w08))
rev(table(i07w08))
rev(table(i08w08))
rev(table(i09w08))
rev(table(i10w08))
rev(table(i11w08))
rev(table(i12w08))
rev(table(i13w08))
rev(table(i14w08))
rev(table(i15w08))
rev(table(i16w08))
rev(table(i17w08))
rev(table(i18w08))
rev(table(i19w08))
rev(table(i20w08))
rev(table(i21w08))
rev(table(i22w08))
rev(table(i23w08))
rev(table(i01w09))
rev(table(i02w09))
rev(table(i03w09))
rev(table(i04w09))
rev(table(i05w09))
rev(table(i06w09))
rev(table(i07w09))
rev(table(i08w09))
rev(table(i09w09))
rev(table(i10w09))

rev(table(i11w09))
rev(table(i12w09))
rev(table(i13w09))
rev(table(i14w09))
rev(table(i15w09))
rev(table(i16w09))
rev(table(i17w09))
rev(table(i18w09))
rev(table(i19w09))
rev(table(i20w09))
rev(table(i21w09))
rev(table(i22w09))
rev(table(i23w09))
rev(table(i01m10))
rev(table(i02m10))
rev(table(i03m10))
rev(table(i04m10))
rev(table(i05m10))
rev(table(i06m10))
rev(table(i07m10))
rev(table(i08m10))
rev(table(i09m10))
rev(table(i10m10))
rev(table(i11m10))
rev(table(i12m10))
rev(table(i13m10))
rev(table(i14m10))
rev(table(i15m10))
rev(table(i16m10))
rev(table(i17m10))
rev(table(i18m10))
rev(table(i19m10))
rev(table(i20m10))
rev(table(i21m10))
rev(table(i22m10))
rev(table(i23m10))
rev(table(i01m11))
rev(table(i02m11))
rev(table(i03m11))
rev(table(i04m11))
rev(table(i05m11))
rev(table(i06m11))
rev(table(i07m11))
rev(table(i08m11))
rev(table(i09m11))
rev(table(i10m11))
rev(table(i11m11))
rev(table(i12m11))
rev(table(i13m11))
rev(table(i14m11))
rev(table(i15m11))
rev(table(i16m11))
rev(table(i17m11))
rev(table(i18m11))
rev(table(i19m11))
rev(table(i20m11))
rev(table(i21m11))

```

rev(table(i22m11))
rev(table(i23m11))

# Code to reproduce MM Table 4:

mitchell.t41 <-
c(i01w06,i02w06,i03w06,i04w06,i05w06,i01w07,i02w07,i03w07,i04w07,i05w07,i01w08,i02w08,i03
w08,i04w08,i05w08,i01w09,i02w09,i03w09,i04w09,i05w09)
martin.t41 <- c(i01m10,i02m10,i03m10,i04m10,i05m10,i01m11,i02m11,i03m11,i04m11,i05m11)
summary(mitchell.t41)
summary(martin.t41)
length(mitchell.t41)
length(martin.t41)
t.test(mitchell.t41,martin.t41,var.equal=TRUE)

mitchell.t42 <-
c(i06w06,i07w06,i08w06,i09w06,i10w06,i06w07,i07w07,i08w07,i09w07,i10w07,i06w08,i07w08,i08
w08,i09w08,i10w08,i06w09,i07w09,i08w09,i09w09,i10w09)
martin.t42 <- c(i06m10,i07m10,i08m10,i09m10,i10m10,i06m11,i07m11,i08m11,i09m11,i10m11)
summary(mitchell.t42)
summary(martin.t42)
length(mitchell.t42)
length(martin.t42)
t.test(mitchell.t42,martin.t42,var.equal=TRUE)

mitchell.t43 <-
c(i11w06,i12w06,i13w06,i14w06,i15w06,i16w06,i17w06,i11w07,i12w07,i13w07,i14w07,i15w07,i16
w07,i17w07,i11w08,i12w08,i13w08,i14w08,i15w08,i16w08,i17w08,i11w09,i12w09,i13w09,i14w09,i
15w09,i16w09,i17w09)
martin.t43 <-
c(i11m10,i12m10,i13m10,i14m10,i15m10,i16m10,i17m10,i11m11,i12m11,i13m11,i14m11,i15m11,i16
m11,i17m11)
summary(mitchell.t43)
summary(martin.t43)
length(mitchell.t43)
length(martin.t43)
t.test(mitchell.t43,martin.t43,var.equal=TRUE)

mitchell.t44 <-
c(i18w06,i19w06,i20w06,i18w07,i19w07,i20w07,i18w08,i19w08,i20w08,i18w09,i19w09,i20w09)
martin.t44 <- c(i18m10,i19m10,i20m10,i18m11,i19m11,i20m11)
summary(mitchell.t44)
summary(martin.t44)
length(mitchell.t44)
length(martin.t44)
t.test(mitchell.t44,martin.t44,var.equal=TRUE)

mitchell.t45 <-
c(i21w06,i22w06,i23w06,i21w07,i22w07,i23w07,i21w08,i22w08,i23w08,i21w09,i22w09,i23w09)
martin.t45 <- c(i21m10,i22m10,i23m10,i21m11,i22m11,i23m11)
summary(mitchell.t45)
summary(martin.t45)
length(mitchell.t45)
length(martin.t45)
t.test(mitchell.t45,martin.t45,var.equal=TRUE)

```

```
# Code for t-tests with no assumption of equal variances, for a Mitchell/Martin comparison for each item:
```

```
# Item 1 [Instructor] (p=0.01, Martin higher)
mitchell.01 <-c(i01w06,i01w07,i01w08,i01w09)
martin.01 <-c(i01m10,i01m11)
c(mean(mitchell.01),sd(mitchell.01),length(mitchell.01))
c(mean(martin.01),sd(martin.01),length(martin.01))
t.test(mitchell.01,martin.01,var.equal=FALSE)
```

```
# Item 2 [Instructor] (p=0.01, Martin higher)
mitchell.02 <-c(i02w06,i02w07,i02w08,i02w09)
martin.02 <-c(i02m10,i02m11)
c(mean(mitchell.02),sd(mitchell.02),length(mitchell.02))
c(mean(martin.02),sd(martin.02),length(martin.02))
t.test(mitchell.02,martin.02,var.equal=FALSE)
```

```
# Item 3 [Instructor] (p=0.02, Martin higher)
mitchell.03 <-c(i03w06,i03w07,i03w08,i03w09)
martin.03 <-c(i03m10,i03m11)
c(mean(mitchell.03),sd(mitchell.03),length(mitchell.03))
c(mean(martin.03),sd(martin.03),length(martin.03))
t.test(mitchell.03,martin.03,var.equal=FALSE)
```

```
# Item 4 [Instructor] (p=0.03, Martin higher)
mitchell.04 <-c(i04w06,i04w07,i04w08,i04w09)
martin.04 <-c(i04m10,i04m11)
c(mean(mitchell.04),sd(mitchell.04),length(mitchell.04))
c(mean(martin.04),sd(martin.04),length(martin.04))
t.test(mitchell.04,martin.04,var.equal=FALSE)
```

```
# Item 5 [Instructor] (p=0.01, Martin higher)
mitchell.05 <-c(i05w06,i05w07,i05w08,i05w09)
martin.05 <-c(i05m10,i05m11)
c(mean(mitchell.05),sd(mitchell.05),length(mitchell.05))
c(mean(martin.05),sd(martin.05),length(martin.05))
t.test(mitchell.05,martin.05,var.equal=FALSE)
```

```
# Item 6 [Instructor/Course] (p=0.01, Martin higher)
mitchell.06 <-c(i06w06,i06w07,i06w08,i06w09)
martin.06 <-c(i06m10,i06m11)
c(mean(mitchell.06),sd(mitchell.06),length(mitchell.06))
c(mean(martin.06),sd(martin.06),length(martin.06))
t.test(mitchell.06,martin.06,var.equal=FALSE)
```

```
# Item 7 [Instructor/Course] (p=0.01, Martin higher)
mitchell.07 <-c(i07w06,i07w07,i07w08,i07w09)
martin.07 <-c(i07m10,i07m11)
c(mean(mitchell.07),sd(mitchell.07),length(mitchell.07))
c(mean(martin.07),sd(martin.07),length(martin.07))
t.test(mitchell.07,martin.07,var.equal=FALSE)
```

```
# Item 8 [Instructor/Course] (p=0.08, Martin higher)
mitchell.08 <-c(i08w06,i08w07,i08w08,i08w09)
martin.08 <-c(i08m10,i08m11)
c(mean(mitchell.08),sd(mitchell.08),length(mitchell.08))
c(mean(martin.08),sd(martin.08),length(martin.08))
```

```

t.test(mitchell.08,martin.08,var.equal=FALSE)

# Item 9 [Instructor/Course] (p=0.02, Martin higher)
mitchell.09 <-c(i09w06,i09w07,i09w08,i09w09)
martin.09   <-c(i09m10,i09m11)
c(mean(mitchell.09),sd(mitchell.09),length(mitchell.09))
c(mean(martin.09),sd(martin.09),length(martin.09))
t.test(mitchell.09,martin.09,var.equal=FALSE)

# Item 10 [Instructor/Course] (p=0.08, Martin higher)
mitchell.10 <-c(i10w06,i10w07,i10w08,i10w09)
martin.10   <-c(i10m10,i10m11)
c(mean(mitchell.10),sd(mitchell.10),length(mitchell.10))
c(mean(martin.10),sd(martin.10),length(martin.10))
t.test(mitchell.10,martin.10,var.equal=FALSE)

# Item 11 [Course] (p=0.02, Martin higher)
mitchell.11 <-c(i11w06,i11w07,i11w08,i11w09)
martin.11   <-c(i11m10,i11m11)
c(mean(mitchell.11),sd(mitchell.11),length(mitchell.11))
c(mean(martin.11),sd(martin.11),length(martin.11))
t.test(mitchell.11,martin.11,var.equal=FALSE)

# Item 12 [Course] (p=0.14, Martin higher)
mitchell.12 <-c(i12w06,i12w07,i12w08,i12w09)
martin.12   <-c(i12m10,i12m11)
c(mean(mitchell.12),sd(mitchell.12),length(mitchell.12))
c(mean(martin.12),sd(martin.12),length(martin.12))
t.test(mitchell.12,martin.12,var.equal=FALSE)

# Item 13 [Course] (p=0.43, Martin higher)
mitchell.13 <-c(i13w06,i13w07,i13w08,i13w09)
martin.13   <-c(i13m10,i13m11)
c(mean(mitchell.13),sd(mitchell.13),length(mitchell.13))
c(mean(martin.13),sd(martin.13),length(martin.13))
t.test(mitchell.13,martin.13,var.equal=FALSE)

# Item 14 [Course] (p=0.04, Martin higher)
mitchell.14 <-c(i14w06,i14w07,i14w08,i14w09)
martin.14   <-c(i14m10,i14m11)
c(mean(mitchell.14),sd(mitchell.14),length(mitchell.14))
c(mean(martin.14),sd(martin.14),length(martin.14))
t.test(mitchell.14,martin.14,var.equal=FALSE)

# Item 15 [Course] (p=0.07, Martin higher)
mitchell.15 <-c(i15w06,i15w07,i15w08,i15w09)
martin.15   <-c(i15m10,i15m11)
c(mean(mitchell.15),sd(mitchell.15),length(mitchell.15))
c(mean(martin.15),sd(martin.15),length(martin.15))
t.test(mitchell.15,martin.15,var.equal=FALSE)

# Item 16 [Course] (p=0.68, Martin higher)
mitchell.16 <-c(i16w06,i16w07,i16w08,i16w09)
martin.16   <-c(i16m10,i16m11)
c(mean(mitchell.16),sd(mitchell.16),length(mitchell.16))
c(mean(martin.16),sd(martin.16),length(martin.16))
t.test(mitchell.16,martin.16,var.equal=FALSE)

```

```

# Item 17 [Course] (p=0.79, Mitchell higher)
mitchell.17 <-c(i17w06,i17w07,i17w08,i17w09)
martin.17   <-c(i17m10,i17m11)
c(mean(mitchell.17),sd(mitchell.17),length(mitchell.17))
c(mean(martin.17),sd(martin.17),length(martin.17))
t.test(mitchell.17,martin.17,var.equal=FALSE)

# Item 18 [Technology] (p=0.62, Martin higher)
mitchell.18 <-c(i18w06,i18w07,i18w08,i18w09)
martin.18   <-c(i18m10,i18m11)
c(mean(mitchell.18),sd(mitchell.18),length(mitchell.18))
c(mean(martin.18),sd(martin.18),length(martin.18))
t.test(mitchell.18,martin.18,var.equal=FALSE)

# Item 19 [Technology] (p=0.049, Martin higher)
mitchell.19 <-c(i19w06,i19w07,i19w08,i19w09)
martin.19   <-c(i19m10,i19m11)
c(mean(mitchell.19),sd(mitchell.19),length(mitchell.19))
c(mean(martin.19),sd(martin.19),length(martin.19))
t.test(mitchell.19,martin.19,var.equal=FALSE)

# Item 20 [Technology] (p=0.35, Martin higher)
mitchell.20 <-c(i20w06,i20w07,i20w08,i20w09)
martin.20   <-c(i20m10,i20m11)
c(mean(mitchell.20),sd(mitchell.20),length(mitchell.20))
c(mean(martin.20),sd(martin.20),length(martin.20))
t.test(mitchell.20,martin.20,var.equal=FALSE)

# Item 21 [Administrative] (p=0.83, Mitchell higher)
mitchell.21 <-c(i21w06,i21w07,i21w08,i21w09)
martin.21   <-c(i21m10,i21m11)
c(mean(mitchell.21),sd(mitchell.21),length(mitchell.21))
c(mean(martin.21),sd(martin.21),length(martin.21))
t.test(mitchell.21,martin.21,var.equal=FALSE)

# Item 22 [Administrative] (p=0.55, Martin higher)
mitchell.22 <-c(i22w06,i22w07,i22w08,i22w09)
martin.22   <-c(i22m10,i22m11)
c(mean(mitchell.22),sd(mitchell.22),length(mitchell.22))
c(mean(martin.22),sd(martin.22),length(martin.22))
t.test(mitchell.22,martin.22,var.equal=FALSE)

# Item 23 [Administrative] (p=0.56, Mitchell higher)
mitchell.23 <-c(i23w06,i23w07,i23w08,i23w09)
martin.23   <-c(i23m10,i23m11)
c(mean(mitchell.23),sd(mitchell.23),length(mitchell.23))
c(mean(martin.23),sd(martin.23),length(martin.23))
t.test(mitchell.23,martin.23,var.equal=FALSE)

```

```

#####
# Standard errors clustered by section #
#####

instructor.female <- c(rep_len(1,length(c(i01w06,i01w07,i01w08,i01w09))),
rep_len(0,length(c(i01m10, i01m11))))

section <- c(rep_len(6,length(i01w06)),rep_len(7,length(i01w07)),
rep_len(8,length(i01w08)),rep_len(9,length(i01w09)),rep_len(10,length(i01m10)),
rep_len(11,length(i01m11)))

total.01 <- c(i01w06,i01w07,i01w08,i01w09,i01m10,i01m11)
total.02 <- c(i02w06,i02w07,i02w08,i02w09,i02m10,i02m11)
total.03 <- c(i03w06,i03w07,i03w08,i03w09,i03m10,i03m11)
total.04 <- c(i04w06,i04w07,i04w08,i04w09,i04m10,i04m11)
total.05 <- c(i05w06,i05w07,i05w08,i05w09,i05m10,i05m11)

total.06 <- c(i06w06,i06w07,i06w08,i06w09,i06m10,i06m11)
total.07 <- c(i07w06,i07w07,i07w08,i07w09,i07m10,i07m11)
total.08 <- c(i08w06,i08w07,i08w08,i08w09,i08m10,i08m11)
total.09 <- c(i09w06,i09w07,i09w08,i09w09,i09m10,i09m11)
total.10 <- c(i10w06,i10w07,i10w08,i10w09,i10m10,i10m11)

total.11 <- c(i11w06,i11w07,i11w08,i11w09,i11m10,i11m11)
total.12 <- c(i12w06,i12w07,i12w08,i12w09,i12m10,i12m11)
total.13 <- c(i13w06,i13w07,i13w08,i13w09,i13m10,i13m11)
total.14 <- c(i14w06,i14w07,i14w08,i14w09,i14m10,i14m11)
total.15 <- c(i15w06,i15w07,i15w08,i15w09,i15m10,i15m11)

total.16 <- c(i16w06,i16w07,i16w08,i16w09,i16m10,i16m11)
total.17 <- c(i17w06,i17w07,i17w08,i17w09,i17m10,i17m11)
total.18 <- c(i18w06,i18w07,i18w08,i18w09,i18m10,i18m11)
total.19 <- c(i19w06,i19w07,i19w08,i19w09,i19m10,i19m11)
total.20 <- c(i20w06,i20w07,i20w08,i20w09,i20m10,i20m11)

total.21 <- c(i21w06,i21w07,i21w08,i21w09,i21m10,i21m11)
total.22 <- c(i22w06,i22w07,i22w08,i22w09,i22m10,i22m11)
total.23 <- c(i23w06,i23w07,i23w08,i23w09,i23m10,i23m11)

table4 <-
data.frame(instructor.female,section,total.01,total.02,total.03,total.04,total.05,total.0
6,total.07,total.08,total.09,total.10,total.11,total.12,total.13,total.14,total.15,total.
16,total.17,total.18,total.19,total.20,total.21,total.22,total.23)

write.table(table4, "e:/table4.txt", sep="\t")

# Import table4.txt into Stata. Then run the following commands:

reg total01 instructorfemale, cluster(section)
reg total02 instructorfemale, cluster(section)
reg total03 instructorfemale, cluster(section)
reg total04 instructorfemale, cluster(section)
reg total05 instructorfemale, cluster(section)
reg total06 instructorfemale, cluster(section)
reg total07 instructorfemale, cluster(section)
reg total08 instructorfemale, cluster(section)
reg total09 instructorfemale, cluster(section)

```

```
reg total10 instructorfemale, cluster(section)
reg total11 instructorfemale, cluster(section)
reg total12 instructorfemale, cluster(section)
reg total13 instructorfemale, cluster(section)
reg total14 instructorfemale, cluster(section)
reg total15 instructorfemale, cluster(section)
reg total16 instructorfemale, cluster(section)
reg total17 instructorfemale, cluster(section)
reg total18 instructorfemale, cluster(section)
reg total19 instructorfemale, cluster(section)
reg total20 instructorfemale, cluster(section)
reg total21 instructorfemale, cluster(section)
reg total22 instructorfemale, cluster(section)
reg total23 instructorfemale, cluster(section)
```

* Stata code for the t-tests

```
ttest total01, by(instructorfemale) unpaired
ttest total02, by(instructorfemale) unpaired
ttest total03, by(instructorfemale) unpaired
ttest total04, by(instructorfemale) unpaired
ttest total05, by(instructorfemale) unpaired
ttest total06, by(instructorfemale) unpaired
ttest total07, by(instructorfemale) unpaired
ttest total08, by(instructorfemale) unpaired
ttest total09, by(instructorfemale) unpaired
ttest total10, by(instructorfemale) unpaired
ttest total11, by(instructorfemale) unpaired
ttest total12, by(instructorfemale) unpaired
ttest total13, by(instructorfemale) unpaired
ttest total14, by(instructorfemale) unpaired
ttest total15, by(instructorfemale) unpaired
ttest total16, by(instructorfemale) unpaired
ttest total17, by(instructorfemale) unpaired
ttest total18, by(instructorfemale) unpaired
ttest total19, by(instructorfemale) unpaired
ttest total20, by(instructorfemale) unpaired
ttest total21, by(instructorfemale) unpaired
ttest total22, by(instructorfemale) unpaired
ttest total23, by(instructorfemale) unpaired
```

```
#####
# Enrollment size as indicated in Texas Tech records #
#####
```

Instructor	Section	Actual N
Mitchell	D6	247
Mitchell	D7	247
Mitchell	D8	243
Mitchell	D9	243
Martin	D10	242
Martin	D11	199

```
Total N for Mitchell in evaluated sections:      980
Total N for Martin in evaluated sections:        441

Mean N for Mitchell in evaluated sections:       245
Mean N for Martin in evaluated sections:        220.5

Total N for Mitchell's evaluations in evaluated sections: 167
Total N for Martin's evaluations in evaluated sections:  51

Response rate in Mitchell's evaluated sections:  17.0%
Response rate in Martin's evaluated sections:    11.6%
```

```
# Student evaluation response rate
n.mitchell  <- 980
n.martin    <- 441
n.evals     <- 167+51
n.simulations <- 10000000
list.prop   <- rep_len(NA,n.simulations)
list.mitchell <- rep_len(NA,n.simulations)
for (i in 1:n.simulations){
  evals      <- c(rep_len(1,n.evals),rep_len(0,n.mitchell+n.martin-n.evals))
  mitchell   <- sample(evals,n.mitchell,replace=FALSE)
  diff.prop  <- 100*(sum(mitchell)/n.mitchell-(n.evals-sum(mitchell))/n.martin)
  list.prop[i]   <- diff.prop
  list.mitchell[i] <- sum(mitchell)
}
stack(100*table(list.prop)/n.simulations)
fraction <- 100*abs(167/980-51/441)
100*length(list.prop[(abs(list.prop)>=fraction)])/n.simulations

# Stata commands for the above:
prtesti 980 0.170408 441 0.115646
prtesti 980 167 441 51, count
```