

Effects of instructor gender and ethnicity on FCQ ratings

October 2016

Perry Sailor, Institutional Research
Office of Data Analytics

Summary of major findings

The study was designed to look for possible effects of instructor gender and ethnicity on FCQ ratings (student ratings of courses and after statistically controlling the effects of class level (graduate vs. undergraduate), size, and department. We examined tenured and tenure-track (TTT) and non-TTT instructors separately, and excluded teaching assistants (TAs). The major findings of the study are:

- Gender differences were exceedingly small; all reported results below are therefore combined across genders.
- The largest effect of instructor ethnicity on ratings was seen among non-TTT Asians (a category which may include both Asian-Americans and non-citizens native to Asia), who were rated 0.22 points lower than whites, about a third of a standard deviation. Other minority non-TTT instructors were rated 0.10 points lower than whites, but this is only about .15 standard deviation units, an exceedingly small difference.
- Among TTT instructors, differences between minority and white instructors were small but discernable, averaging 0.14 rating points for both Asians and other minorities, or .29 standard deviation units.
- When the population studied is restricted to instructors who taught at least three sections in the three-year period studied, some small but observable differences mostly remained.
- These findings are similar to those of a study we did 14 years ago.

Detailed procedure and results

Procedure

We looked at FCQ ratings for all terms from three academic years – 2013-14, 2014-15, and 2015-16.

Each observation on the data file (N=20,745) represented an instructor/section combination, with a mean rating from students in the section on FCQ items 7 (global course rating) and 8 (global instructor rating). The ratings were restricted to those for

- Lectures, recitations, labs, and seminars
- Instructor groups TTT (tenured and tenure-track) and OTH (other primary instructors, not TTT). Group TA (teaching assistants) was excluded.

We statistically removed from each rating the effects of class size, level (undergrad vs. grad), and individual department. This was done by entering these variables as predictors into a

regression modeling program¹ and obtaining an expected rating based on these factors alone. This expected rating was then subtracted from the actual rating, yielding a residual. After converting each *section* mean rating to a residual in this fashion, a single mean residual for each *instructor* was then calculated, by averaging the instructor's ratings across all sections taught.

The study included ratings on FCQ item 7, course ratings, as well as item 8, instructor ratings. However, because course ratings and instructor ratings were so highly correlated ($r=.87$), the remainder of this report will discuss only instructor ratings. All statements and differences reported below that apply to one apply to the other also.

Results

All results below are shown in terms of residual scores. The mean of all the residual scores across the entire population is zero, by definition (see technical note below). This means that a residual of, for example, 0.1 can be interpreted as "0.1 points above average, after adjusting for class size, level, and department." The magnitude of a mean residual, as with any mean score, can be evaluated by comparing it to the standard deviation.

Table 1.1-1.3 Residual ratings controlling for class size, class level, and department, all instructors fall 2013 through summer 2016.

Table 1.1 Tenured/tenure track (TTT)

Ethnic group	N	Mean	S. D.
Asian	146	-.12	.57
Other minority	113	-.12	.48
White/unknown	980	.02	.47
All	1,239	-.01	.49

Table 1.2 Other instructors (OTH)

Ethnic group	N	Mean	S. D.
Asian	143	-.30	.73
Other minority	191	-.18	.71
White/unknown	1,978	-.08	.64
All	2,312	-.10	.66

Table 1.3 All – TTT and OTH

Ethnic group	N	Mean	S. D.
Asian	289	-.21	.66
Other minority	304	-.16	.64
White/unknown	2,958	-.04	.59
All	3,551	-.07	.61

Among TTT instructors, there was little difference between other ethnic groups compared to whites, with the -.12 lower rating for Asians and other minorities amounting to about .24 standard deviation units ($.12/.49 = .24$), which is fairly small. However, among non-TTT instructors, the difference between ratings of Asians and whites was a bit larger – the mean rating for Asians was .22 below that for whites, a difference of about a third of a standard deviation.

¹ SAS procedure GLM (General Linear Model).

A separate analysis was done after eliminating from the population instructors who taught only 1 or 2 sections across the 9 terms. Restricting the analysis to instructors who taught at least three sections sharply attenuated the large negative difference between Asian non-TTT instructors and other groups, and also resulted in a substantial drop in standard deviation for that group, indicating that most of the negative effect seen in the above table was due to a few instructors who taught only one or two sections each and received extremely low ratings. Perhaps the low ratings they received is the reason they only taught one or two sections – their departments realized they were ineffective instructors and gave them no more teaching assignments. This is just speculation, however. These results are summarized in Table 2 below.

Table 2-2.3 Residual ratings controlling for class size, class level, and department, instructors with at least 3 sections taught from fall 2013 through summer 2016.

Table 2.1 Tenured/tenure track (TTT)

Ethnic group	N	Mean	S. D.
Asian	132	-.11	.49
Other minority	99	-.14	.48
White/unknown	832	.03	.44
All	1,063	-.00	.45

Table 2.2 Other instructors (OTH)

Ethnic group	N	Mean	S. D.
Asian	72	-.17	.53
Other minority	93	-.05	.53
White/unknown	1,113	-.01	.52
All	1,278	-.02	.52

Table 2.3 All – TTT and OTH

Ethnic group	N	Mean	S. D.
Asian	204	-.13	.50
Other minority	192	-.10	.51
White/unknown	1,945	.01	.49
All	2,341	-.01	.49

Technical note:

Because the residual values were calculated on the individual section mean ratings, before the reduction to one mean score per instructor, the overall mean across *instructors*, collapsed across sections, will not necessarily be 0; in fact, in this dataset it is -.07 for instructor ratings, -.02 for course ratings.