# Faculty Quality Evaluations Report

Seyoung Lee

November 30, 2023

## Introduction

There are many possible ways to assess the quality of an instructor's teaching, such as third-party websites that provides ratings or university-specific course assessments conducted by provincial faculty. While these evaluations may offer a degree of reliability, they may be susceptible to biases such as perceptions of class difficulty or gender biases, which can influence the final assessments of instructors and even promotions.

The Vice Provost for Faculty at the University of Southern North Dakota in Hoople is working to assess these biases effectively, aiming to identify the variables that most affects the ratings of teaching quality. This effort not only ackowledges excellence among instructors but also acknowledges potential biases that studnets may have when evaluating the quality of their professors.

In this study, we examine a dataset comprising of 366 professors at USND, each having received over 10 ratings on an instructor evaluation website. Using variables, including gender, department, attractiveness, helpfulness, and other factors, our objective with this dataset is to address the following research questions:

1. Are instructors' quality ratings associated with (1) Gender, (2) Perceptions of attractivess, (3) Easiness of classes, or (4) Discipline of the instructor?
2. Is the relationship between easiness and instructor quality dependent on instructor gender and discipline?
3. What model yields the best prediction model for instructor quality ratings by students among easiness, gender and discipline of instructors?

As a brief statement of results, we can conclude that discipline cannot accurately determine instructor quality, while gender, attractiveness perceptions, and easiness can. The relationship between easiness and instructor quality, however, is not dependent on instructor gender and discipline. Given these results, we can assume that easiness and gender as predictor models are most likely to predict quality among easiness, gender, and discipline.

## Exploratory Data Analysis & Data Summary

### Overall data information

There are 366 observations (n=366), each corresponding to a professor with a minimum of 10 ratings spanning several years. The dataset encompasses 17 variables, including both factor variables and ratings on 5-point scales. These variables offer average ratings and additional characteristics

about the instructors. Nonetheless, for the purposes of addressing the research questions, only the following five variables will be considered:

quality: average quality rating (Between 1, worst, to 5, best)

gender: instructor gender (factor - female/male)

pepper: perceptions of attractiveness (factor - yes/no)

discipline: discipline instructor teaches (factor with levels - Hum/SocSci/STEM/Pre-prof)

easiness: average easiness rating (Between 1, worst, to 5, best)

**Univariate data**

Evaluating the distribution of each relevant variable prior to examining the connections between them is crucial for the reliability of the representation of the data and reliability of the associations in the model.

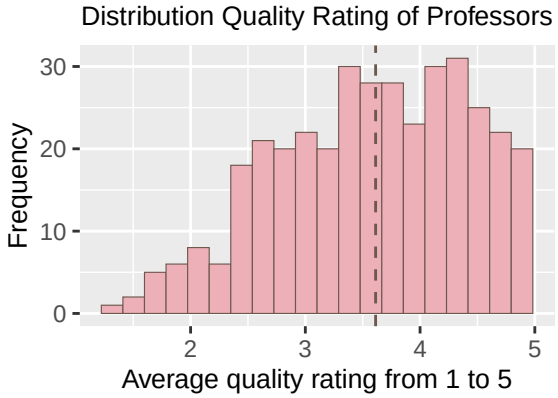


Figure 1: histogram for the distribution of instructors' quality ratings.

Figure 1 illustrates the distribution of professor ratings provided by instructors. The data reveals a noticeably high median, indicating a left skew in the distribution. This means that students are more likely to give all instructors a positive rating; this may affect the overall effect size of the variables, since we know that the ratings are inflated it toward better ratings.

Table 1: ratio of instructors female to male

| Gender | Frequency |
|---|---|
| female | 43.443 |
| male | 56.557 |

Table 2: student ratio on perceptions attractiveness

| y/n | Frequency |
|---|---|
| no | 87.432 |
| yes | 12.568 |

Table 1 and 2 show the ratio between groups in the data: male/female or perceived attractive/unattractive. We see that the frequency of female to male is somewhat equal with a ratio of 43 to 56. However, the frequency between considered unattractive to attractive by students is skewed with a ratio of 87 to 12. It is important to note that this may affect the data as the representation of instructors in either group is not proportional in the final mode.

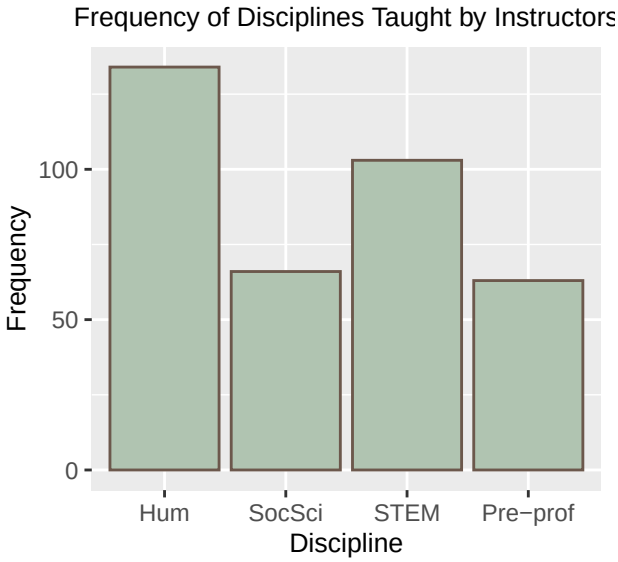


Figure 2: bar graph depicting the distribution of disciplii

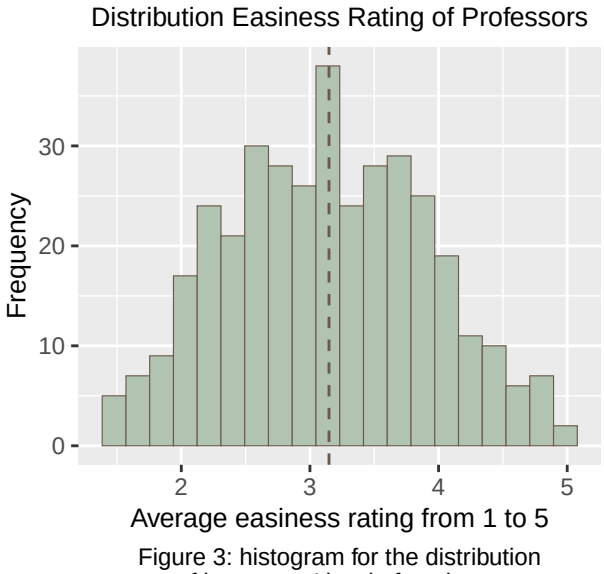


Figure 3: histogram for the distribution of instructors' level of easiness

Figure 2 demonstrates the categorical distribution between disciplines represented in the instructors. We see that while humanities is represented the most, Social sciences and pre-professional instructors are the least represented. This may indicate that the prediction model represents this specific distribution of data.

Figure 3 displays the distribution of rate of easiness in professors. This data exhibits a more even distribution, resembling a normal distribution, in contrast to the distribution of average quality rating.

### Bivariate relationships

To accurately predict instructor quality, it is important to explore the individual bivariate relationships between each individual variable to instructor quality. This assesses which variables might have an immediate affect on the response variable or correlations between two predictor variables.

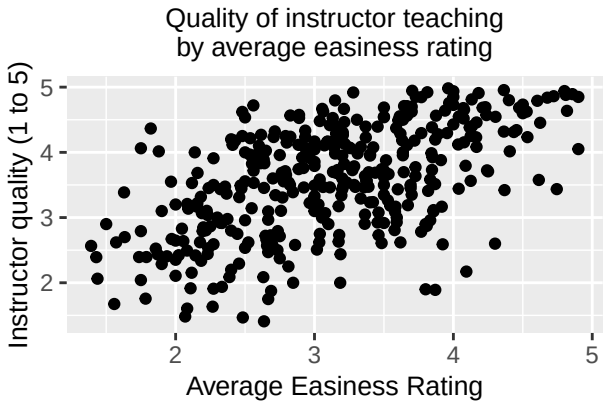


Figure 4: scatterplot of quality by easiness

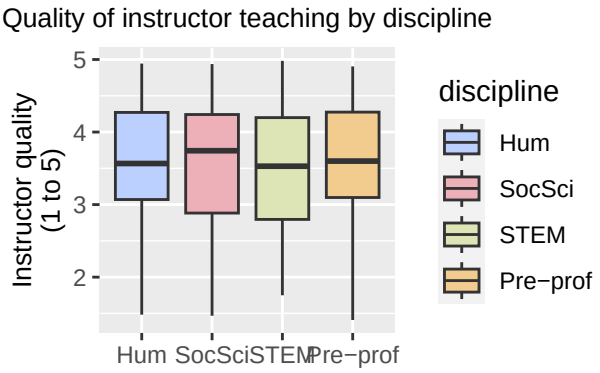


Figure 5: bar plot of quality by discipline

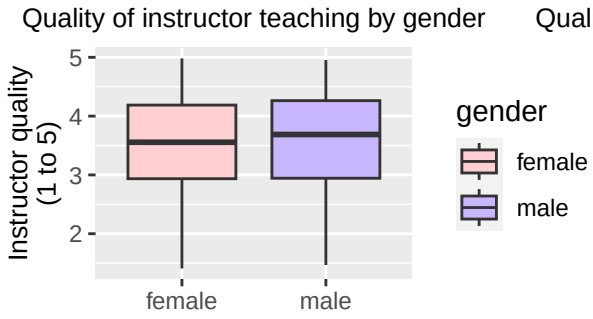


Figure 6: barplot of quality by gender

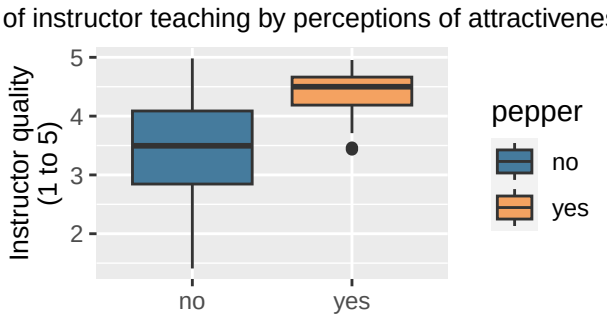


Figure 7: barplot of quality by attractiveness

Figure 4, 5, 6, and 7 each represent the individual relationships between a predictor variable with the quality of instructor teaching. We see that there is a slight positive correlation with average easiness and instructor quality, but a large correlation between perceptions of attractiveness and the distribution of instructor quality.

### Limitations

It is important to note that there are some limitations to the data after performing exploratory data analysis to the instructor dataset. For example, the univariate distribution of the quality of instructor teaching showed that it was skewed left, while the distribution of instructor attractiveness rating was also skewed, which may misrepresent the predictive quality of the model we test.

## Methods

**1. Are instructors' quality ratings associated with (1) Gender, (2) Attractiveness according to students, (3) Easiness of classes, or (4) Discipline of the instructor?**

To address this question, we examine the significance of these associations individually by analyzing simple linear regression p-values. Predictors with p-values smaller than alpha=0.05 may signify that the variable is significant. This is shown in the Table 3 and further interpreted through its effect sizes of the individual variables. We use this table as a display of which independent variables are significantly associated with quality.

**2. Is the relationship between easiness and instructor quality dependent on instructor gender and discipline?**

To test whether this relationship is dependent on gender and discipline, we use the model that suggests that this association and dependencies exist and assess the significance of the effect sizes. By testing the following model, we will be able to determine whether there is a relationship between easiness and instructor quality if there is a dependency on gender and discipline.

$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 I(X_2 = gendermale) + \beta_3 II(X_2 = disc1) + \beta_4 II(X_2 = disc2) + \beta_5 II(X_2 = disc3) + \beta_6 III(X_4 = easiness : gendermale) + \beta_7 IV(X_5 = easiness : disc1) + \beta_8 IV(X_5 = easiness : disc2) + \beta_9 IV(X_5 = easiness : disc3)$

We perform a global F-test to show that the model with interactions yields a significant difrferent in slop of easiness between gender and discipline factors (which implies dependencies).

Using Analysis of Variance, we compare this model with interactions with the following model with no interactions.

$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 I(X_2 = gendermale) + \beta_3 II(X_3 = disc1) + \beta_4 II(X_3 = disc2) + \beta_5 II(X_3 = disc3)$

**3. What model yields the best prediction model for instructor quality ratings by students among easiness, gender and discipline of instructors?**

To find which combination of final dependencies and independent associations between the predictors and instructor quality, we will use the previous model tests and assumptions, such as the F-test and Analysis of Variance. Given these statistics, we found that the final model that best predicts instructor quality ratings might be the following:

$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 X_2(gendermale)$

To test this model, we checked the significance values for each estimate in this model, including its interpretation of the effect size along with its confidence interval.

We also provided the Akaike Information Criterion (AIC), which measures strengths of predictions of the model, to show the model's capability in predicting instructor quality.

From the results, we gathered that this model can indeed adequately predict instructor quality rating. However,

**Model diagnostics**

Before testing both these models, we also perform basic residual diagnostics to make sure it adequately captured the information in this data:

1. The mean of residuals is approximately zero: From both Figure 7 and 8, we see that both models that we will test have a mean of approximately zero, according to the fitted blue dashed line.

2. The residuals have a constant variance: Both figures additionally show that the residuals have a constant variance with no deformities in the shape of the fitted values against the residuals.

3. The residuals are approximately normal: We can assume that the residuals for both linear functions are approximately normal from the Normal QQ plots (Figure 10 and 11), as the data shows very little diversion from the fitted red line.

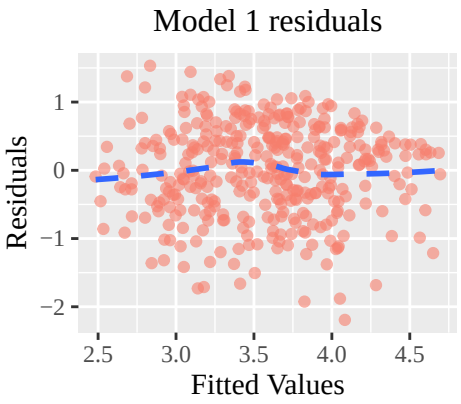


Figure 8: residuals for model with all variables and interaction terms

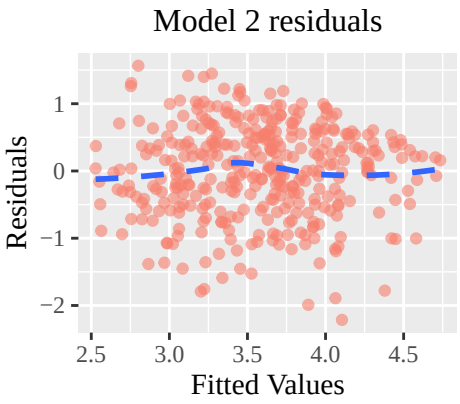


Figure 9: residuals for model with reduced interaction terms

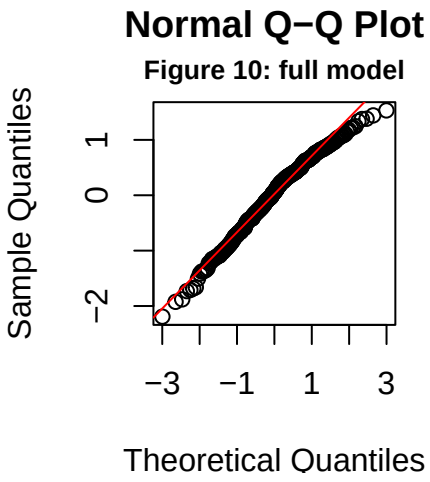


**Figure 10: full model**

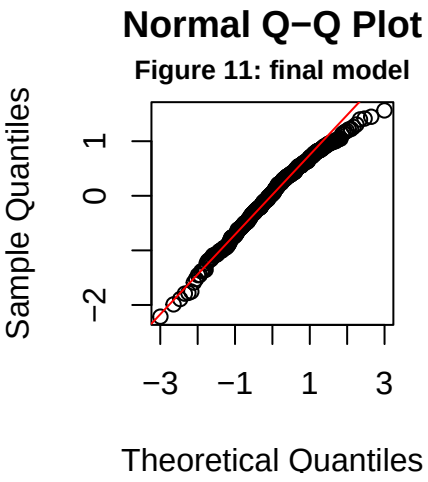


**Figure 11: final model**

## Results

**1. Are instructors' quality ratings associated with (1) Gender, (2) Attractiveness according to students, (3) Easiness of classes, or (4) Discipline of the instructor?**

The following table displays the individual significance values for each predictor. In observing the table, we see that the predictors related to discipline and gender lack statistical significance, while easiness and attractiveness do. This may be important to consider when creating a final model that predicts instructor quality, as we are aware which factors in the predictor values are independently better at predicting quality.

Table 3: Individual significance values per predictor

| Predictor | P_values |
|---|---|
| Easiness | 2.87e-32 |
| Discipline Math | 1.45e-162 |
| Discipline Social Science | 0.809 |
| Discipline STEM | 0.656 |
| Discipline Pre-prof | 0.729 |
| Gender Female | 1.23e-173 |
| Gender Male | 0.386 |
| Attractive (No) | 3.73e-232 |
| Attractive (Yes) | 1.24e-13 |

**2. Is the relationship between easiness and instructor quality dependent on instructor gender and discipline?**

To show that there is an interaction, or if the new coefficients are all zero or not, we use an F-test to compare two models.

Without interaction:

$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 I(X_2 = gendermale) + \beta_3 II(X_3 = disc1) + \beta_4 II(X_3 = disc2) + \beta_5 II(X_3 = disc3)$

With interaction:

$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 I(X_2 = gendermale) + \beta_3 II(X_2 = disc1) + \beta_4 II(X_2 = disc2) + \beta_5 II(X_2 = disc3) + \beta_6 III(X_4 = easiness : gendermale) + \beta_7 IV(X_5 = easiness : disc1) + \beta_8 IV(X_5 = easiness : disc2) + \beta_9 IV(X_5 = easiness : disc3)$

In this context, we are testing whether there is a significant difference in slope of easiness between gender and discipline factors. This implies that we are testing for an interaction, or if the new coefficients are all zero or not using the F-test:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$

$H_a : \beta_i \neq 0$ for at least one $i$

The following ANOVA table represents the significance of the interactions by comparing the two models:

Table 4: ANOVA table without interactions / with interactions

| model | df.residual | statistic | p.value |
|---|---|---|---|
| Model without interactions | 360 | NA | NA |
| Model with interactions | 356 | 0.2523539 | 0.9081518 |

From the results of the ANOVA test, we fail to reject the null hypothesis that at least one coefficient from this subset of independent variables had a significant effect on instructor quality (F(4, 356),

p=0.908). Hence, the slope for easiness is not significantly different between gender factors and discipline factors, which means that the relationship between easiness and instructor quality is not dependent on instructor gender and discipline.

**3. What model yields the best prediction model for instructor quality ratings by students given easiness, gender and discipline of instructors?**

Using the step function on the full model,

$$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 I(X_2 = gendermale) + \beta_3 II(X_2 = disc1) + \beta_4 II(X_2 = disc2) + \beta_5 II(X_2 = disc3) + \beta_6 III(X_4 = easiness : gendermale) + \beta_7 IV(X_5 = easiness : disc1) + \beta_8 IV(X_5 = easiness : disc2) + \beta_9 IV(X_5 = easiness : disc3) ,$$

the results show that a significant model predicting quality reduced to the following model,

$$Y = \beta_0 + \beta_1 X_1(easiness) + \beta_2 X_2(gendermale)$$

We can visualize the model likeso,


Figure 11: linear model fit of quality given easiness

To assess this model, we first perform individual t-tests to interpret the effect size of each predictor variable, including its factors, and to see whether this effect size is significant.

Table 5: Model summary

| | Estimate | P-Value | Lower CI | Upper CI |
|---|---|---|---|---|
| (Intercept) | 1.5474506 | 4.93e-20 | 1.2346825 | 1.8602187 |
| easiness | 0.6175675 | 4.38e-33 | 0.5261317 | 0.7090034 |
| gendermale | 0.1614395 | 0.0273 | 0.0181742 | 0.3047048 |

Based on the results from the model, we can assume the following in the estimates in our model:

(Intercept), or genderfemale baseline: Holding all other predictors constant, the expected value in quality ratings for female instructors when easiness rating is 0 is 1.547, on average (95% CI[1.235, 1.860]) with its significant estimate (t(366), p=4.93e-20).

easiness: On average, holding all other predictors constant, an additional unit rating in easiness is associated with 0.617 unit increase in instructor quality for female instructors (95% CI [0.526, 0.709]), given the estimate's significance (t(366), p=4.38e-33).

gendermale: On average, an additional unit increase in easiness rating in male instructors is associated with a 0.1614 increase in rating of instructor quality compared to female instructors (95% CI[0.0181, 0.3047]), given the estimate's significance (t(366), p=0.0273)

An alternative approach to gauging the strength of the relationship is by examining the Akaike Information Criterion (AIC). With a lower AIC signifying a better prediction model fitness, this model with an AIC of 3.00 offers more accurate predictions of instructor quality compared to the model featuring easiness, gender, and discipline interactions as predictors, which has an AIC of 10.00.

Considering its notably low AIC value and the statistical significance of its predictor variables, we can assert that easiness and gender can reliably predict instructor quality.

## Discussion

To finalize this research, we provide some conclusions to the presented research questions according to the results.

1. Are instructors' quality ratings associated with any of the following predictors (1) Gender, (2) Attractiveness according to students, (3) Easiness of classes, or (4) Discipline?

Given the independent simple linear regression models and significance testing, we can assume that gender, attractiveness according to students, and easiness of classes of instructors is significantly associated with instructor quality.

2. Is the relationship between easiness and instructor quality dependent on instructor gender and discipline?

No, the relationship between easiness and instructor quality is not dependent on instructor gender or discipline.

3. What model yields the best prediction model for instructor quality ratings by students among easiness, gender and discipline of instructors?

Among these variables, we can assume that easiness and gender are better variables for predicting instructor quality.

### Limitations and improvements

Before reaching conclusions, it is important to note any limitations in the data and analysis as well as improvements for later research. Firstly, the data is observational rather than experimental, so we cannot assume causation. For instance, we are not sure if students are more likely to think

an instructor is attractive due to their teaching quality or if students are more likely to think an instructor is better at teaching due to their perceptions of attractiveness.

Another limitation in this analysis is that we did not take the other 17 variables into account. Doing independent associations with those variables on top of the variables we accounted for may give insight into more qualities that might predict instructor quality ratings.

In future research, it may also be worth gathering data on comments students make on instructors to further gauge a positive or negative sentiment in addition to gathering a more comprehensive dataset from randomly selected colleges to represent the overall population, rather than a single school.

Given the results, conclusions, and limitations, some recommendations we can make to improve the course evaluations is to also provide an area where students can comment on professors as well as a further analysis on prediction of instructor quality with more variables.