Examining How Raters and Various Factors Influence the Success of a New General Education Program

Alana Willis
Department of Statistics and Data Science
Carnegie Mellon University
alanaw@andrew.cmu.edu

November 29, 2021

Abstract

Many colleges and universities use a curriculum known as General Education to allow students to broaden their course of study. The Dietrich College at Carnegie Mellon University is in the process of implementing a new General Education program for undergraduates and we were tasked to measure the success of this program. We examined data on 91 artifacts produced by students that were rated by three raters from three different departments. We performed exploratory data analysis, calculated ICCs and percent exact agreements to examine differences in ratings between different rubrics and raters, and we used mixed effect models to determine what factors influenced ratings. We found that raters often differed in their interpretations of rubrics and in turn, highly affected the fairness of ratings among the artifacts. We also found that other factors such as Semester and Sex had an influence on ratings. We feel that the college should consider a training class for the raters so that their interpretation and uses of the rubrics will be more consistent and the Dean's Office may want to consider broadening the scope of the project to include more courses and departments within the College.

1 Introduction

College is a time for students to learn and experience new things in order to expand their range of knowledge in many subjects, as well as advance knowledge in a specific subject of interest. Many colleges and universities use a curriculum known as General Education to allow students to broaden their course of study. The Dietrich College at Carnegie Mellon University (2021) says their General Education curriculum is "designed to help [students] maintain and enhance [their] intellectual breadth in ways that are more closely tailored to [their] particular interests...and enhance [their] intellectual growth by creating stimulating comparisons and synergies from disparate fields of study."

However, the college is in the process of implementing a new General Education program for undergraduates. The new program specifies a set of courses and experiences that all undergraduates must take. In order to determine whether the new program is successful, the college hopes to rate the quality of student work in the General Education courses each year based on seven rubrics, using raters from across the college. The most recent General Education course of interest is Freshman Statistics. With the data collected from the raters, we hope to be able to answer the following questions with our analysis:

> *(1)* Is the distribution of ratings for each rubric pretty much indistinguishable from the other rubrics, or are there rubrics that tend to get especially high or low ratings? Is the distribution of ratings given by each rater pretty much indistinguishable from the other raters, or are there raters that tend to give especially high or low ratings?
>
> *(2)* For each rubric, do the raters generally agree on their scores? If not, is there one rater who disagrees with the others? Or do they all disagree?
>
> *(3)* More generally, how are the various factors in this experiement (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?
>
> *(4)* Is there anything else interesting to say about this data?

It is important for the General Education curriculum to be as innovative and inclusive as possible to prepare students to tackle complex problems in work and in life. The Dietrich College at Carnegie Mellon University specifically wants their students to be able to "challenge social, political, and global concerns such as inequality and injustice, climate change and voting." We hope the findings of this study influence positive change in the General Education curriculum.

2 Data

The data is composed of 91 project papers that were randomly sampled from a Fall and Spring section of Freshman Statistics class at Carnegie Mellon University. The individual papers will be referred to as "artifacts" throughout this analysis. Three raters from three different departments across the college were asked to rate these artifacts on the basis of seven rubrics, described in Table 1. The rating scale for all rubrics is shown in Table 2.

Short Name	Full Name	Description
RsrchQ	Research Question	Given a scenario, the student generates, critiques or evaluates a relevant empirical research question.
CritDes	Critique Design	Given an empirical research question, the student critiques or evaluates to what extent a study design convincingly answer that question.
InitEDA	Initial EDA	Given a data set, the student appropriately describes the data and provides initial Exploratory Data Analysis.
SelMeth	Select Method(s)	Given a data set and a research question, the student selects appropriate method(s) to analyze the data.
InterpRes	Interpret Results	The student appropriately interprets the results of the selected method(s).
VisOrg	Visual Organization	The student communicates in an organized, coherent and effective fashion with visual elements (charts, graphs, tables, etc.).
TxtOrg	Text Organization	The student communicates in an organized, coherent and effective fashion with text elements (words, sentences, paragraphs, section and subsection titles, etc.).

Table 1: Descriptions of the rating rubrics.

Rating	Meaning
1	Student does not generate any relevant evidence.
2	Student generates evidence with significant flaws.
3	Student generates competent evidence; no flaws, or only minor ones.
4	Student generates outstanding evidence; comprehensive and sophisticated.

Table 2: Rating scale for each rubric.

The raters did not know which class or which students produced the artifacts that they rated. Thirteen of the 91 artifacts were rated by all three raters and the remaining 78 artifacts were rated by only one rater. The variables available for analysis are defined in Table 3.

Variable Name	Values	Description
(X)	1, 2, 3, . . .	Row number in the data set
Rater	1, 2 or 3	Which of the three raters gave a rating
(Sample)	1, 2, 3, . . .	Sample number
(Overlap)	1, 2, . . . , 13	Unique identifier for artifact seen by all 3 raters
Semester	Fall or Spring	Which semester the artifact came from
Sex	M or F	Sex or gender of student who created the artifact
RsrchQ	1, 2, 3 or 4	Rating on Research Question
CritDes	1, 2, 3 or 4	Rating on Critique Design
InitEDA	1, 2, 3 or 4	Rating on Initial EDA
SelMeth	1, 2, 3 or 4	Rating on Select Method(s)
InterpRes	1, 2, 3 or 4	Rating on Interpret Results
VisOrg	1, 2, 3 or 4	Rating on Visual Organization
TxtOrg	1, 2, 3 or 4	Rating on Text Organization
Artifact	(text labels)	Unique identifier for each artifact
Repeated	0 or 1	1 = this is one of the 13 artifacts seen by all 3 raters

Table 3: Variable descriptions. Variable names in () were not used in any analysis.

The file ratings.csv contains data organized exactly as in Table 3. The file tall.csv contains the same data, but organized so that each row contains a single rating and the rubric for that rating is listed in the column labeled Rubric.

Below in figure 1, the distribution of ratings is shown for the artifacts seen by all three raters. A rating of 2 was the most common, while a rating of 4 was the least common. The same distribution can be seen in Appendix 1, where we constructed the same histogram but for the full dataset. In this histogram the distribution of ratings is consistent with a rating of 2 being the most used and a rating of 4 being the least.

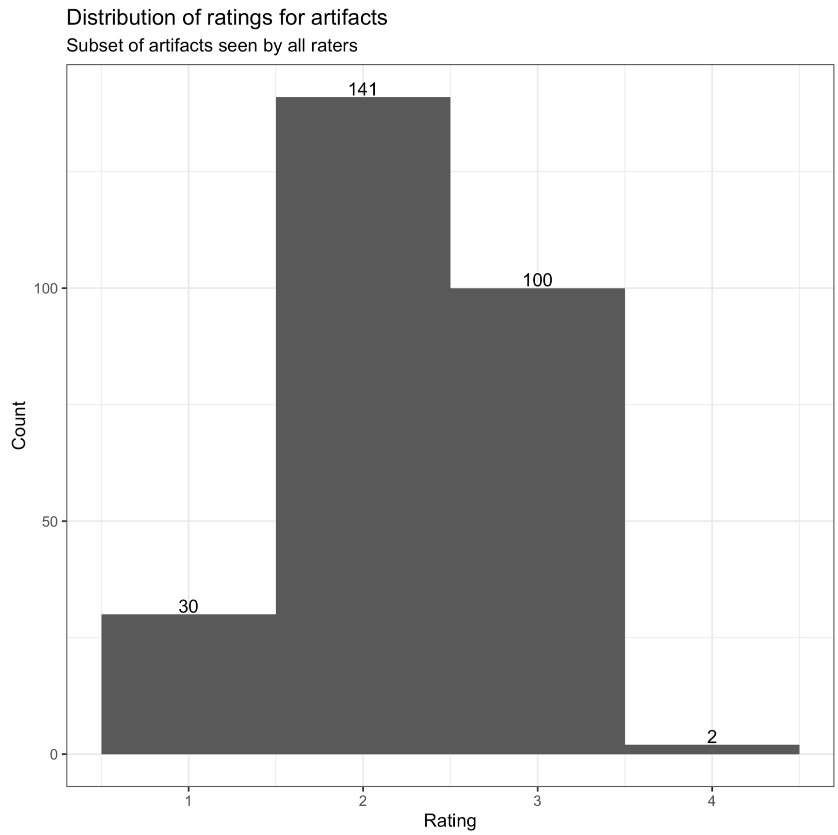


Figure 1: Distribution of ratings for artifacts seen by all three raters.

	Min.	1st Q.	Median	Mean	3rd Q.	Max	SD
RsrchQ	1	2	2	2.35	3	4	.59
CritDes	1	1	2	1.87	3	4	.84
InitEDA	1	2	2	2.44	3	4	.70
SelMeth	1	2	2	2.07	2	3	.49
InterpRes	1	2	3	2.49	3	4	.61
VisOrg	1	2	2	2.41	3	4	.67
TxtOrg	1	2	3	2.60	3	4	.70

Table 4: Summary Statistics of ratings by rubric on all 91 artifacts.

Table 4 shows summary statistics of the ratings by each rubric and table 5 shows summary statistics of ratings by each rater. We will discuss the differences of ratings between rubrics and raters later in this report.

	Min	1st Q.	Median	Mean	3rd Q.	Max	SD
Rater 1	1	2	2	2.35	3	4	.70
Rater 2	1	2	2	2.43	3	4	.70
Rater 3	1	2	2	2.18	3	4	.69

Table 5: Summary Statistics of ratings by rater on all 91 artifacts.

One thing to note about the full data is missingness within some of the variables. There is a missing value for Sex and a missing Rating value in the CritDes and VisOrg rubrics caused by Rater 1 and Rater 2. It is important to note that in any modeling that we do where Rating is the outcome variable, R will automatically drop the two observations. This means that the full data sets vary slightly for models involving CritDes where we would be missing a rating from Rater 2 and for models involving VisOrg we would be missing a rating from Rater 1. We will also have to be careful of the missing Sex value. For the purposes of this analysis we have decided to keep the missing value as a third category for Sex. We do not want the entry to be dropped when modeling and there is not a sufficient way to recode it within the other two categories. The code for discovering this missingness can be found in Appendix 1. There is no problem of missing entries in the subset data.

3 Methods

All statistical modeling and visualizations were made using the R language and environment for statistical computing (R Core Team, 2017).

We started by loading the tall.csv and rating.csv datasets in our R environment. We then created a modified data set of the tall.csv, that is a subset of the thirteen artifacts seen by all three raters. We checked for missingness in the full data set and created tables to see exactly where these missing values were present (Appendix 1). We also calculated summary statistics of important variables in the data (Appendix 1).

3.1 Distribution of Ratings for Rubrics and Raters

After these initial steps, we performed exploratory data analysis to compare the distribution of ratings overall, the distribution of ratings for each of the rubrics, and the distribution of ratings among the raters; helping us to answer the first research question. This was done for both the subset of data and the full data set to determine whether the thirteen artifacts are representative of the whole set. All of the EDA can be found in Appendix 1.

3.2 Measuring Agreement Between Raters

To begin to address the second research question, we wanted to measure the intraclass correlation (ICC) between raters. The ICC's help us determine whether the raters are generally in agreement or not regarding each rubric. To do this we used the subset data for just the thirteen artifacts seen by all three raters and modified it more to only include the specific rubric of interest. We then fit seven random-intercept models, one for each rubric, and manually calculated the seven ICC's based on the summary output.

This leads to the next portion where we calculated the percent exact agreement between pairs of raters. This will help to determine who is agreeing with whom on each rubric since the ICC cannot tell us which raters might be contributing to disagreement. To do this we made 2-way tables of counts for the ratings for each pair of raters on each rubric and divided the counts on the main diagonal by the total number of counts.

We repeated the process of calculating ICC's for the full data set to see if they agree with the ICCs we calculated for subset data. We are not able to do the same for the percent exact agreements because not all raters rated every artifact. All processes for this section can be found in Appendix 2.

3.3 Various Factors Relation to Ratings

Next we want to characterize how the other variables in the data set interact with ratings, leading us to answer research question three. We performed manual variable selection models on each of the seven rubrics for the full data set. For fixed effect models we used AIC/BIC and likelihood ratio tests to compare models to find the best one for each rubric. We also calculated the ICC's again based on these new models. After finding the best fixed effect model for each rubric we tested some different random effects and interactions to see if they led to an improved model. We used AIC/BIC to decide on the best models. This method is flawed because it doesn't let you directly examine interactions with Rubric. To combat this we used the full data set without subsetting by each rubric individually. We again manually selected fixed effects, random effects, and considered interactions to find the best model. We used likelihood ratio tests and AIC/BIC to compare models. All related code to this process can be found in Appendix 3.

3.4 Additional Exploratory Data Analysis

Finally to address research question four and to complete the analysis, we performed some additional EDA using Semester and Sex to gather any other conclusions we felt weren't covered by the final models.

4 Results

4.1 Distribution of Ratings for Rubrics and Raters

In figure 1 we were able to visualize the distribution of ratings amongst all artifacts in both the subset of data and the full dataset. The distribution was the same for both datasets giving us some initial indication that the subset is a good representation of the full data. Figure 2 displays the distribution of ratings for each rubric for only the artifacts seen by all three raters. Listed below are some conclusions based on Figure 2:

- Rubrics InitEDA, RsrchQ, SelMeth, and VisOrg all present similar patterns between ratings. Each of them have rating 2 as the most commonly given and rating 4 the least commonly given, which matches the overall distribution of ratings.
- Rubric CritDes seems to be the most harshly rated with most of the raters giving a 1 or 2 rating.
- The rubrics InterpRes and TxtOrg seem to be the most highly rated with the majority of raters giving either a 2 or a 3 rating. The InterpRes and TxtOrg rubrics are also the only ones to have a 4 rating.

We constructed the same histogram for the full dataset, which can be seen in Appendix 1. The distribution of ratings among the rubrics is the same for the full dataset, giving us even more confirmation that the subset is a good representation of the full data.

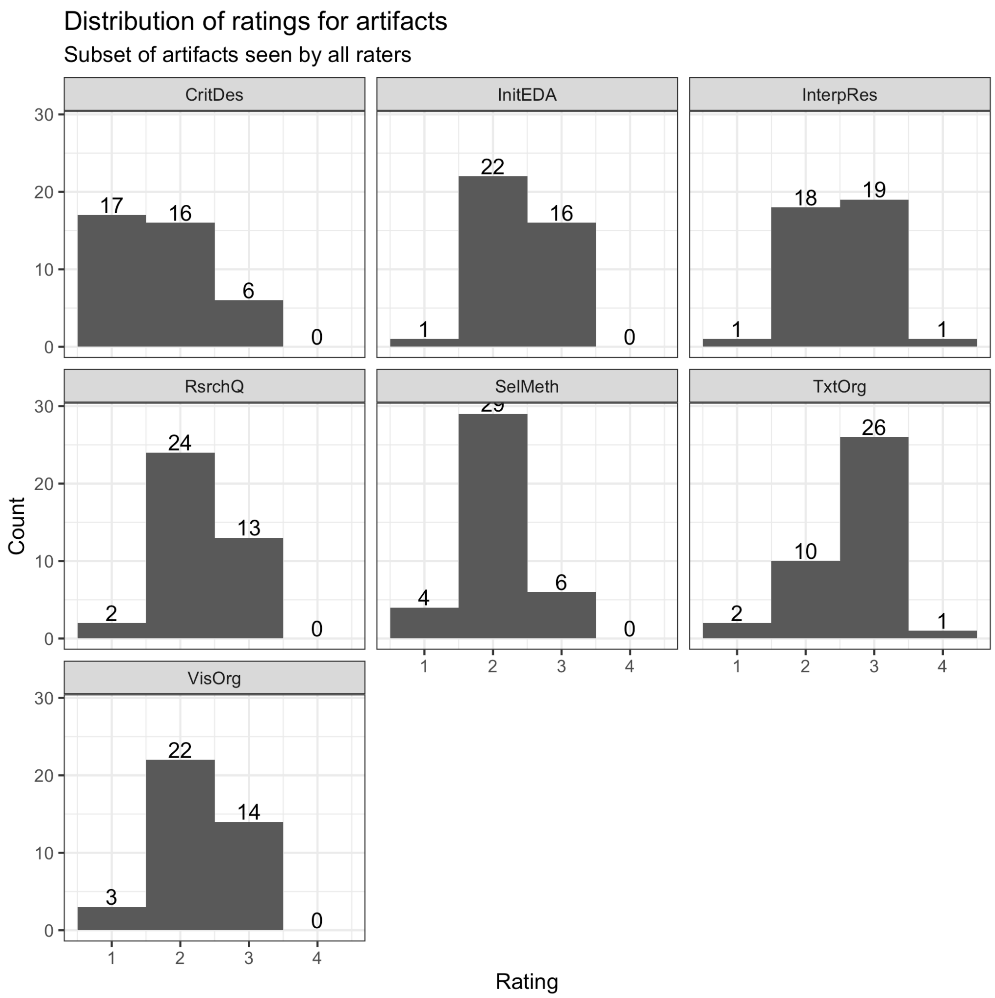


Figure 2: Distribution of ratings by rubric for subset of artifacts seen by all raters.

Lastly, we wanted to compare the distributions of ratings among the three raters. The distribution for the subset data can be seen in Figure 3 and the full data in Appendix 1. Listed below are some conclusions based on Figure 3 and the histogram of the full data in Appendix 1:

- Figure 3 shows that the distributions of ratings for each rater is mostly similar, with 3 being the most common and 4 being the least common. This also matches the overall distribution of ratings. However, there is a slight difference when we visualize the distribution for the full data set.
- In these histograms it can be seen that Rater 2 gives ratings of 2 and 3 about the same amount and drastically more than they give a 1 or 4. Rater 2 is also the most generous, giving the most 3 and 4 ratings.
- Rater 3 seems to be the harshest, giving the most 1 and 2 ratings.

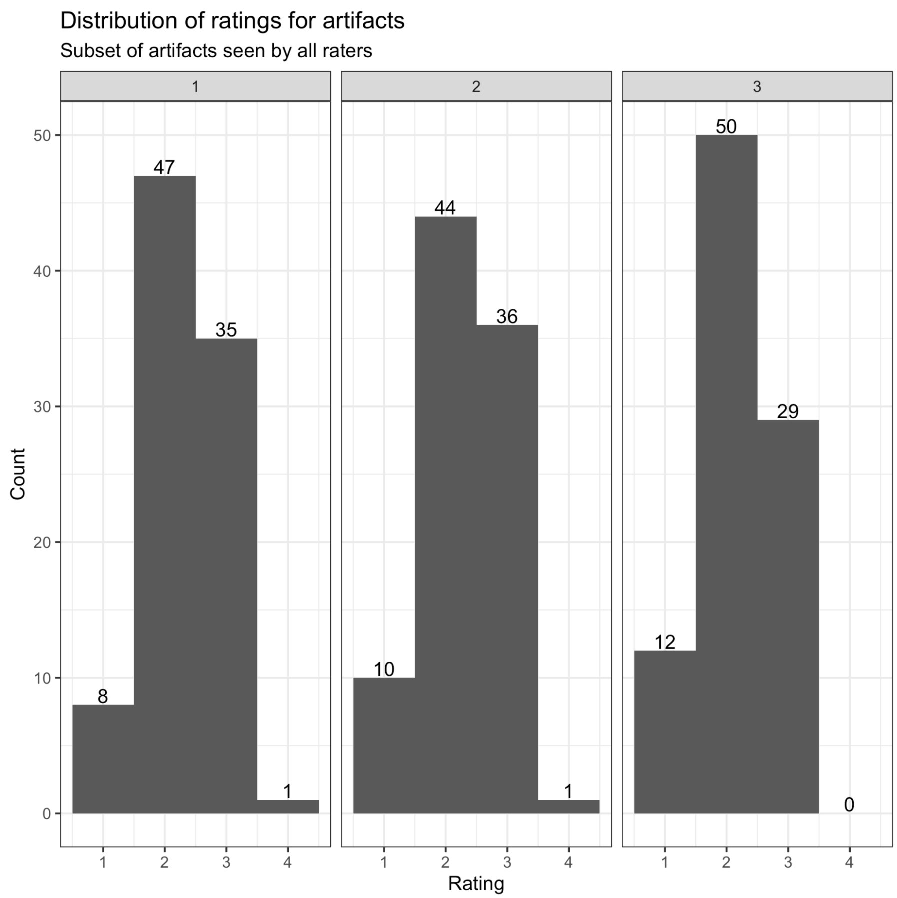


Figure 3: Distribution of ratings between raters for artifacts seen by all raters.

4.2 Measuring Agreement Between Raters

The next aspect we wanted to characterize was the agreement or disagreement between raters. The first measure of agreement among the raters we considered was the Intraclass Correlation (ICC) which represents the common correlation among the raters' ratings for each artifact. To calculate it, we treated each artifact as a cluster of three ratings and fit seven random-intercept models, one for each rubric, using the subset data. The fitted models and actual ICC calculations can be found in Appendix 2.

In table 6 below, the final subset ICC's for each rubric were recorded. A high ICC represents a high correlation among the rater while a low ICC represents low correlation among the raters. Below are some conclusions based on the 2nd column of Table 6:

- The raters seem to agree most on rubric VisOrg with an ICC of .59 and the least on rubric TxtOrg with an ICC of .14.
- Rubrics RsrchQ and InterpRes also have pretty low agreement with ICC's of .19 and .23, respectively.
- The raters agree moderately with each other on the remaining rubrics with the ICC's falling between .49 and .57.

Rubric	**ICC (subset)**	**ICC (full)**	**ICC (after variable selection)**
RsrchQ	0.189	0.210	0.207
CritDes	0.573	0.673	0.671
InitEDA	0.493	0.687	0.688
SelMeth	0.500	0.472	0.453
InterpRes	0.230	0.220	0.198
VisOrg	0.592	0.661	0.665
TxtOrg	0.143	0.188	0.191

Table 6: Intraclass correlations between raters

The ICC's are good at categorizing overall agreement between raters but they cannot tell us which raters might be contributing to disagreement. To determine this we calculated the percent exact agreement between pairs of raters. A table of the percent exact agreements can be found in table 7 below. From the table it can be concluded that:

- Raters 1 and 2 have the most variation in agreement among the rubrics with the highest agreement on SelMeth at 92% and the lowest on RsrchQ at only 38%.
- Raters 1 and 3 had moderately consistent agreement across all rubrics, with percent exact agreement ranging from 54% to 77%.
- Similarly, raters 2 and 3 have moderate agreement on most rubrics falling between 54% and 77%. However, they have high agreement on the InitEDA rubric at 85%.

	Rater 1 & Rater 2	Rater 1 & Rater 3	Rater 2 & Rater 3
RsrchQ	38%	77%	54%
CritDes	54%	62%	54%
InitEDA	69%	54%	85%
SelMeth	92%	62%	69%
InterpRes	62%	54%	62%
VisOrg	54%	77%	77%`
TxtOrg	69%	62%	54%

Table 7: Percent Exact Agreement among raters.

To complete the analysis for this research question, we repeated the earlier ICC calculations but for the full data set to compare the agreement between raters further. These models and calculations can be found in Appendix 2. The ICC's for the full dataset can be found in table 6. From this table we can see that:

- The ICC's increase for most rubrics except for SelMeth and InterpRes, which decrease slightly.
- The InitEDA rubric had the highest change with an ICC increase going from around .49 with the subset data to about .69 with the full data.

4.3 Various Factors Relation to Ratings

To begin to address the third research question, we performed manual variable selection on the seven random intercept models for each rubric. First, adding in only fixed effects, then interactions, and finally random effects. The process of variable selection for the seven models can be found in Appendix 3. Each of the final models were chosen using likelihood ratio tests and AIC/BIC. The final models with the coefficient outputs for each rubric can be seen below in table 8. The conclusions made from Table 8 can be seen below:

- Rubrics RsrchQ, CritDes, InitEDA, and TxtOrg all stayed with the random intercept model as the best model.
- The SelMeth rubric gained new fixed effects with Rater and Semester with each rater increasing the rating by about 2 points and the artifacts from the Spring semester decreased ratings by less than half a point than those from the Fall semester.
- The InterpRes rubric gained Rater as a fixed effect with each rater increasing rating by a little more than 2 points. The same applies to the final model for the VisOrg rubric.
- The Repeated and Sex variables were not selected in the final model for any of the rubrics.

	RsrchQ	CritDes	InitEDA	SelMeth	InterpRes	VisOrg	TxtOrg
Fixed Effects							
Intercept	2.35	1.90	2.44	---	---	---	2.59
Rater 1	---	---	---	2.25	2.70	2.38	---
Rater 2	---	---	---	2.23	2.59	2.65	---
Rater 3	---	---	---	2.23	2.14	2.28	---
SemesterS19	---	---	---	-0.36	---	---	---
Sex	---	---	---	---	---	---	---
Repeated	---	---	---	---	---	---	---
Rubric	---	---	---	---	---	---	---
Random Effects							
Artifact (τ^2)	0.07	0.49	0.37	0.09	0.06	0.29	0.09
Residual (σ^2)	0.28	0.24	0.17	0.11	0.25	0.15	0.40

Table 8: Rubric final models and coefficient estimates.

After finding the final models, we compared the ICC's to those found with the subset of data and the full data set before variable selection. All ICC results can be seen in table 6 and the calculations in Appendix 3. The ICC's after variable selection only varied slightly from those in the models prior, some only by hundredths of a point.

While this approach is great for determining how other factors influence each rubric specifically, it does not allow for direct examination of interactions of the variable Rubric, since each model considers only one rubric at a time. To combat this problem we constructed a new null model that adds the entire Rubric variable as a random effect, grouped by Artifact.

We performed manual variable selection by adding fixed effects, testing interactions, and trying random effects. The process can be seen in Appendix 3.

- The final model selected Rater, Semester, and Rubric as fixed effects along with the interaction between Rater and Rubric. It also selected Rubric and Rater as random effects. The coefficient outputs can be found in Appendix 3.
- All of the fixed effects are significant in the model and majority of the interaction terms are as well.
- All raters increase ratings, with rater 2 having the most influence.
- The rubrics also have a positive association with rating.
- Semester decreases ratings as well as all of the interactions between Rater and Rubric.

4.4 Additional Exploratory Data Analysis

We felt that the models did not fully explain some of the variation of ratings between the sex of the students and the semester in which the artifact came from, therefore we decided to do some exploratory data analysis on those two variables in particular using both the full data and the subset data. The distribution of ratings by semester can be found for the subset data and for the full data in Appendix 4. Looking at the raw count distributions, we can see that the Fall semester has significantly more artifacts than the spring semester. For both semesters a rating of 2 is given the most and a rating of 4 is given the least. This led us to calculating and visualizing the proportion of ratings on the full data set for each semester to combat the significant difference in samples for each semester. From figure 4, we can see that:

- There is only a slight difference in the amount of times a rating of 2, 3, or 4 is given between semesters.
- A rating of 1 is more commonly given in the Spring semester than in the Fall semester.
- The overall distribution for the proportion of ratings is the same for both semesters.

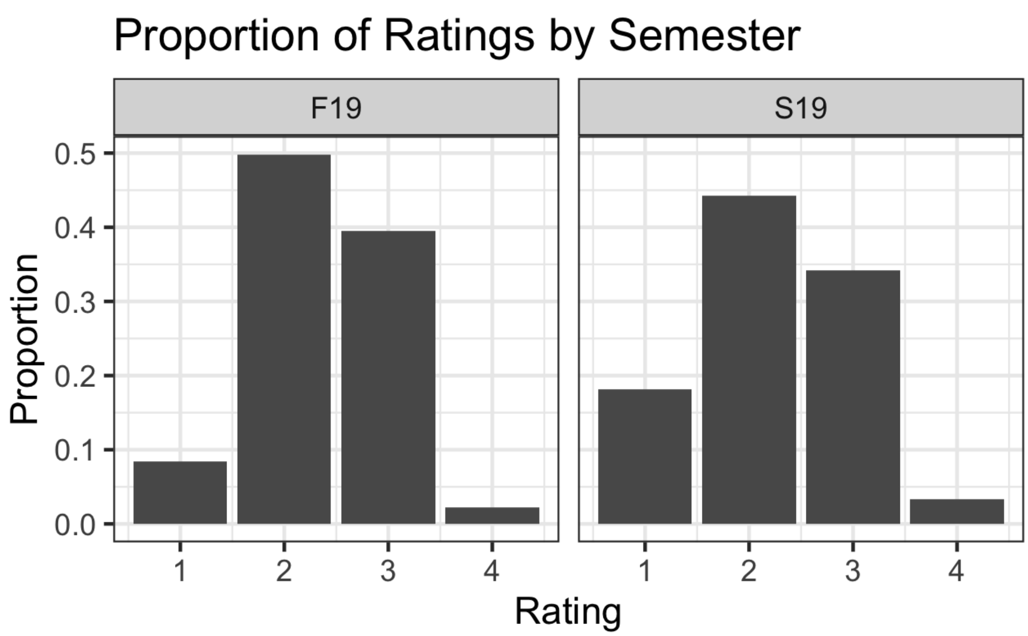


Figure 4: Proportion of rating by semester.

The distribution for ratings by Male and Female for the subset data can be seen in figure 5 and the histogram for the full data can be seen in Appendix 4. The distribution between the sexes is similar with a rating of 2 being given the most and a rating of 4 given the least. The distribution is also the same when done for the full dataset.

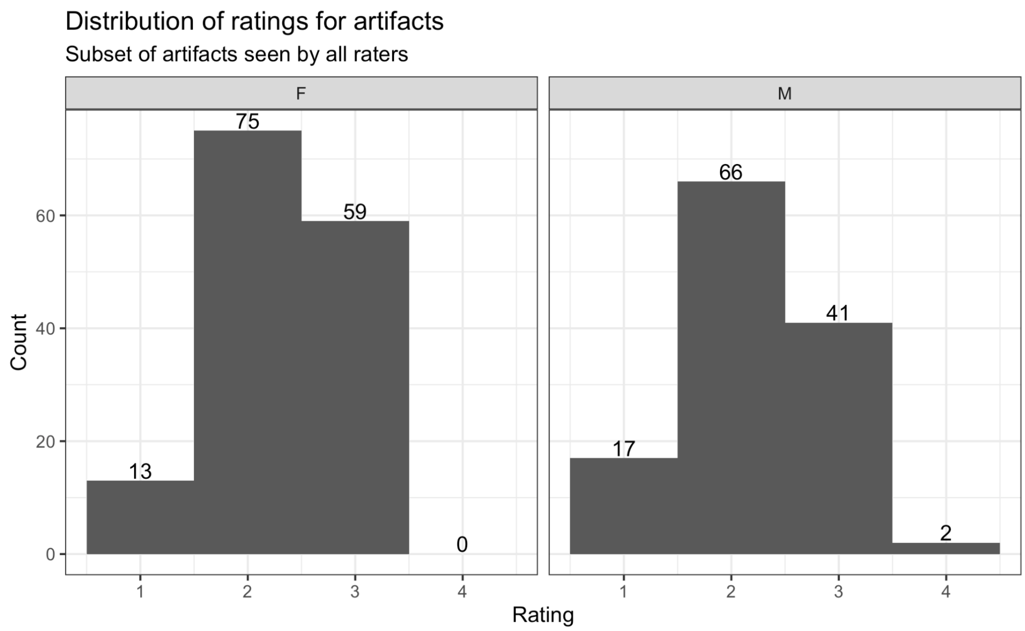


Figure 5: Distribution of ratings by sex.

We also wanted to visualize the differences between ratings by each rater and each of the rubrics (Figure 6). It is clear that there are differences in the ways the raters rated the seven rubrics. For example, in the InterpRes rubric, Rater 3 seemed to grade a bit more harshly than Raters 1 and 2.

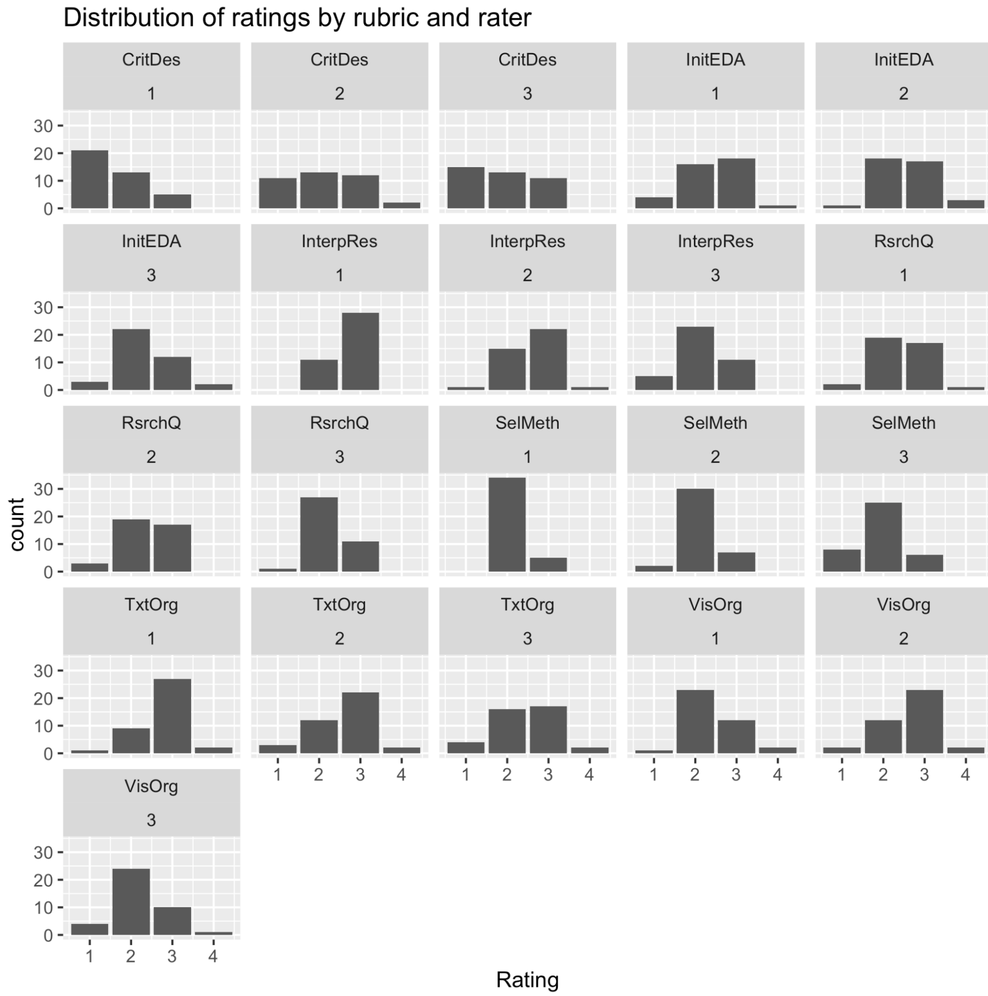


Figure 6: Distribution of ratings by rubric and rater.

5 Discussion

In this study, we were examining four main concerns related to the ratings of three raters of student work, known as artifacts, based on seven rubrics from a Freshman Statistics general education course in hopes of creating a new general education curriculum for Dietrich College at Carnegie Mellon University. We found that Rater 2 seemed to be the most generous rater, while Rater 3 was the harshest. Raters 1 and 2 disagree the most, while Raters 1 and 3 were more consistently in agreement. The SelMeth rubric has the most influence from other factors such as Semester and Rater. We also found that the Raters do have a lot of influence on the ratings of each artifact and that this influence differs based on the rubric. From these main conclusions we feel that the Dean's Office should consider a training course for the raters. This may decrease the individual influence from each rater and lead to a more fair evaluation of the rubrics. We will now discuss the conclusions of the research questions in further detail.

5.1 Distribution of Ratings for Rubrics and Raters

First, we wanted to characterize the distribution of ratings for each of the rubrics to see if there were rubrics that got especially high or low ratings. We also wanted to visually compare the ratings between the raters. We found that the CritDes rubric got the lowest overall ratings while InterpRes and TxtOrg were the most highly rated rubrics. As far as the individual raters it we found that Rater 2 is the most generous giving mostly ratings of 2 or 3 and the most 4 ratings while Rater 3 was the harshest giving mostly ratings of 1 and 2. This difference may suggest that the raters need some kind of training to make their ratings more consistent across the board. We also concluded that the subset data for only the artifacts seen by all raters was a good representation of the full dataset.

5.2 Measuring Agreement Between Raters

Second, we wanted to know if the raters were agreeing or disagreeing with each other. Based on ICC calculations of the random intercept models for the subset data we concluded that the raters agreed the most on the VisOrg rubric with an ICC of .59 and they disagreed on the TxtOrg rubric with an ICC of only .19. These variations give us some initial indication that the raters are rating differently and may have some underlying biases. While the ICC's are a good measure of agreement between the raters we also wanted to see exactly which raters are contributing. We calculated the percent exact agreements for each pair of raters. From this we concluded that Raters 1 and 2 have a lot of variation in their agreement. They can rate very similar at 95% or very different at 38%. Raters 1 and 3 were consistently in agreement more than 50% of the time for all rubrics and had the lowest variability between each rubrics. Raters 2 and 3 fall in the middle of the other pairs. These findings are even more indication that there are biases in the raters. At this point Rater 2 seems to be causing much of the differences in ratings. Next, we compared the ICC's for the full dataset to those from the subset data. We found that there were small changes in the ICC's, but nothing too drastic. We mostly attribute the change to a larger sample size as we went from considering only 13 artifacts to considering all 91 artifacts.

4.3 Various Factors Relation to Ratings

Third, we wanted to determine how various factors such as Rater, Semester, Sex, Repeated, and Rubric related to the ratings. We found the best model for each of the seven rubrics independently and we found the best model for all of the rubrics combined. For the rubrics RsrchQ, CritDes, InitEDA and TxtOrg we found that the random intercept model was the best. This means that it is possible no outside factors are affecting the ratings for these rubrics. The SelMeth rubric gained new fixed effects with Rater and Semester while, VisOrg and InterpRes gained Rater as a fixed effect. The introduction of fixed effects in the models indicates that the raters have some underlying differences for these specific rubrics influencing their ratings as well as the Semester specifically for the SelMeth rubric. The Repeated and Sex variables were not selected in the final model for any of the rubrics. This indicates that there is no change in rating for the rubrics regardless of students gender and ratings between rubrics don't seem to change drastically between the artifacts seen by all raters and the ones seen by only one rater. We compared ICC's one last time after the variable selection process and found that they only varied slightly for each rubric and in most cases only by hundredths of a point. For the final combined model we found that Rater, Semester, and Rubric were fixed effects along with the interaction between Rater and Rubric. It also selected Rubric and Rater as random effects. Based on Rater being a random effect of the final model we can determine that each rater's rating on a specific artifact differs by some random variation depending on the artifact. The Rater and Rubric interaction leads us to believe that each rater uses the rubrics in a way that is not similar to the way other rater's use the rubrics. Rubric being random effect tells us that there are different average scores on each rubric, but the rubric averages vary a bit from one artifact to the next, by a small random effect that depends on the artifact. These interactions further suggest that the raters should be trained more, to make the raters ratings more similar to each other.

4.4 Additional Exploratory Data Analysis

Lastly, we wanted to consider how ratings varied by Sex and Semester since these variables were not included in many of the models. We found that the Fall semester has significantly more artifacts than the Spring semester. A rating of 1 is more commonly given in the Spring semester than in the Fall semester however, the overall distribution for the proportion of ratings is the same for both semesters. These factors may be the reason Semester becomes a significant fixed effect in the final model and the Spring decreases ratings when compared to the Fall. There does not seem to be a significant difference in ratings between the sexes which is good news in terms of bias from the raters. We also wanted to visualize the differences between ratings by each rater and each of the rubrics. Figure 4 only further proved that there are differences in the ways the raters are using the rubrics.

This is by far not a perfect analysis. One weakness of this study is that it is not representative of the entire Dietrich College. We have only analyzed data from a single course in a single department. We do not believe these results could be translated to other courses and departments. A future study could include at least one course from each department in the College so that the success of the new general education curriculum can be measured on more information. Overall, we feel that the college should consider a training class for the raters so that their interpretation and uses of the rubrics will be more consistent before they truly make a decision on the new general education curriculum. The Office may also want to consider broadening the scope of the project to include more courses and departments within the College.

References

Dietrich College (2021), *General Education Program for Students Enrolled Prior to Fall 2021*. Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh PA. Accessed Nov 28, 2021 at https://www.cmu.edu/dietrich/gened/previous-curriculum/index.html

Junker, B. W. (2021). *Project 02 assignment sheet and data for 36-617: Applied Regression Analysis.* Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh PA. Accessed Nov 25, 2021 from https://canvas.cmu.edu/courses/25337/files/folder/Project02

R Core Team (2017), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

RStudio Team (2020), *R Studio: Integrated Development Environment for R*. RStudio, PBC, Boston MA. Available at http://www.rstudio.com/.

Sheather, S.J. (2009), *A Modern Approach to Regression with R*. New York. Springer Science and Business Media LLC.

Technical Appendix

Alana Willis

11/13/2021

Contents


```
library(tidyverse)
library(arm)
library(lme4)
library(plyr)
library(ggplot2)
library(vtable)
library(kableExtra)
```

Appendix 1. Initial Data/Library Imports & Data Exploration

To begin, we loaded the original full data set and created a modified data set that is a subset of just the artifacts seen by all three raters. We checked for missingness and did some summary statistics for the important variables. After this we began EDA.

```
tall.full <- read.csv("~/Desktop/Fall_21/Applied Linear Models/Project_2/tall.csv")
ratings <- read.csv("~/Desktop/Fall_21/Applied Linear Models/Project_2/ratings.csv")

#Subset of data for 13 rubrics seen by all raters
tall <- tall.full %>%
```

```r
  filter(Repeated == "1") %>%
  dplyr::select(-c(X))

#Tables of rating counts to show where NA's are located
tmp0 <- lapply(split(as.factor(tall.full$Rating),tall.full$Rubric),summary)
tmp <- data.frame(matrix(0,nrow=5,ncol=7)) ## seven rubrics...
names(tmp) <- names(tmp0)
row.names(tmp) <- c(paste("Rating",1:4),"<NA>")
for (i in names(tmp0)) {
tmp[,i] <- tmp[,i] + c(tmp0[[i]],0)[1:5]
}
tmp
```

```
##          CritDes InitEDA InterpRes RsrchQ SelMeth TxtOrg VisOrg
## Rating 1      47       8         6      6      10      8      7
## Rating 2      39      56        49     65      89     37     59
## Rating 3      28      47        61     45      18     66     45
## Rating 4       2       6         1      1       0      6      5
## <NA>           1       0         0      0       0      0      1
```

```r
tmp0 <- lapply(split(as.factor(tall.full$Rating),tall.full$Rater),summary)
tmp <- data.frame(matrix(0,nrow=5,ncol=3)) ## three raters...
names(tmp) <- names(tmp0)
row.names(tmp) <- c(paste("Rating",1:4),"<NA>")
for (i in names(tmp0)) {
tmp[,i] <- tmp[,i] + c(tmp0[[i]],0)[1:5]
}
names(tmp) <- paste("Rater",1:3)
tmp
```

```
##          Rater 1 Rater 2 Rater 3
## Rating 1      29      23      40
## Rating 2     125     119     150
## Rating 3     112     120      78
## Rating 4       6      10       5
## <NA>           1       1       0
```

```r
# Can see which rater/rubrics line up in terms of missing values
tall.full[apply(tall.full,1,function(x){any(is.na(x))}),]
```

```
##       X Rater Artifact Repeated Semester Sex  Rubric Rating
## 161 161     2       45        0      S19   F CritDes     NA
## 684 684     1      100        0      F19   F  VisOrg     NA
```

```r
# Can see the missing value for sex here.
head(tall.full)
```

```
##   X Rater Artifact Repeated Semester Sex Rubric Rating
## 1 1     3       05        1      F19   M RsrchQ      3
## 2 2     3       07        1      F19   F RsrchQ      3
## 3 3     3       09        1      S19   F RsrchQ      2
```

```
## 4 4     3       08        1      S19   M RsrchQ      2
## 5 5     3        5        0      F19     RsrchQ      3
## 6 6     3        6        0      F19   M RsrchQ      2
```

```r
# Make the "tall" be consistent with the "ratings" coding.
tall.full$Sex[nchar(tall.full$Sex)==0] <- "--"

#Summary stats for rubrics and raters
summary(ratings)
```

```
##        X           Rater        Sample          Overlap       Semester
##  Min.   :  1   Min.   :1   Min.   :  1.00   Min.   : 1   Length:117
##  1st Qu.: 30   1st Qu.:1   1st Qu.: 31.00   1st Qu.: 4   Class :character
##  Median : 59   Median :2   Median : 60.00   Median : 7   Mode  :character
##  Mean   : 59   Mean   :2   Mean   : 59.89   Mean   : 7
##  3rd Qu.: 88   3rd Qu.:3   3rd Qu.: 89.00   3rd Qu.:10
##  Max.   :117   Max.   :3   Max.   :118.00   Max.   :13
##                                             NA's   :78
##      Sex                RsrchQ        CritDes         InitEDA
##  Length:117         Min.   :1.00   Min.   :1.000   Min.   :1.000
##  Class :character   1st Qu.:2.00   1st Qu.:1.000   1st Qu.:2.000
##  Mode  :character   Median :2.00   Median :2.000   Median :2.000
##                     Mean   :2.35   Mean   :1.871   Mean   :2.436
##                     3rd Qu.:3.00   3rd Qu.:3.000   3rd Qu.:3.000
##                     Max.   :4.00   Max.   :4.000   Max.   :4.000
##                                    NA's   :1
##     SelMeth        InterpRes         VisOrg          TxtOrg
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :3.000   Median :2.000   Median :3.000
##  Mean   :2.068   Mean   :2.487   Mean   :2.414   Mean   :2.598
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :3.000   Max.   :4.000   Max.   :4.000   Max.   :4.000
##                                  NA's   :1
##    Artifact            Repeated
##  Length:117         Min.   :0.0000
##  Class :character   1st Qu.:0.0000
##  Mode  :character   Median :0.0000
##                     Mean   :0.3333
##                     3rd Qu.:1.0000
##                     Max.   :1.0000
##
```

```r
sd(ratings$RsrchQ)
```

```
## [1] 0.5918446
```

```r
sd(ratings$CritDes, na.rm=T)
```

```
## [1] 0.8395669
```

```
sd(ratings$InitEDA)
```

```
## [1] 0.6995641
```

```
sd(ratings$SelMeth)
```

```
## [1] 0.486481
```

```
sd(ratings$InterpRes)
```

```
## [1] 0.6104744
```

```
sd(ratings$VisOrg, na.rm=T)
```

```
## [1] 0.67333
```

```
sd(ratings$TxtOrg)
```

```
## [1] 0.6955503
```

```
Rater.1 <- tall.full %>%
  mutate(Rater = as.factor(Rater)) %>%
  filter(Rater == "1") %>%
  dplyr::select(c(Rater, Rating))
summary(Rater.1)
```

```
##  Rater       Rating
##  1:273   Min.   :1.000
##  2:  0   1st Qu.:2.000
##  3:  0   Median :2.000
##          Mean   :2.349
##          3rd Qu.:3.000
##          Max.   :4.000
##          NA's   :1
```

```
sd(Rater.1$Rating, na.rm=T)
```

```
## [1] 0.6974383
```

```
Rater.2 <- tall.full %>%
  mutate(Rater = as.factor(Rater)) %>%
  filter(Rater == "2") %>%
  dplyr::select(c(Rater, Rating))
summary(Rater.2)
```

```
##  Rater       Rating
##  1:  0   Min.   :1.00
##  2:273   1st Qu.:2.00
##  3:  0   Median :2.00
##           Mean   :2.43
##           3rd Qu.:3.00
##           Max.   :4.00
##           NA's   :1
```

```r
sd(Rater.2$Rating, na.rm=T)
```

```
## [1] 0.699691
```

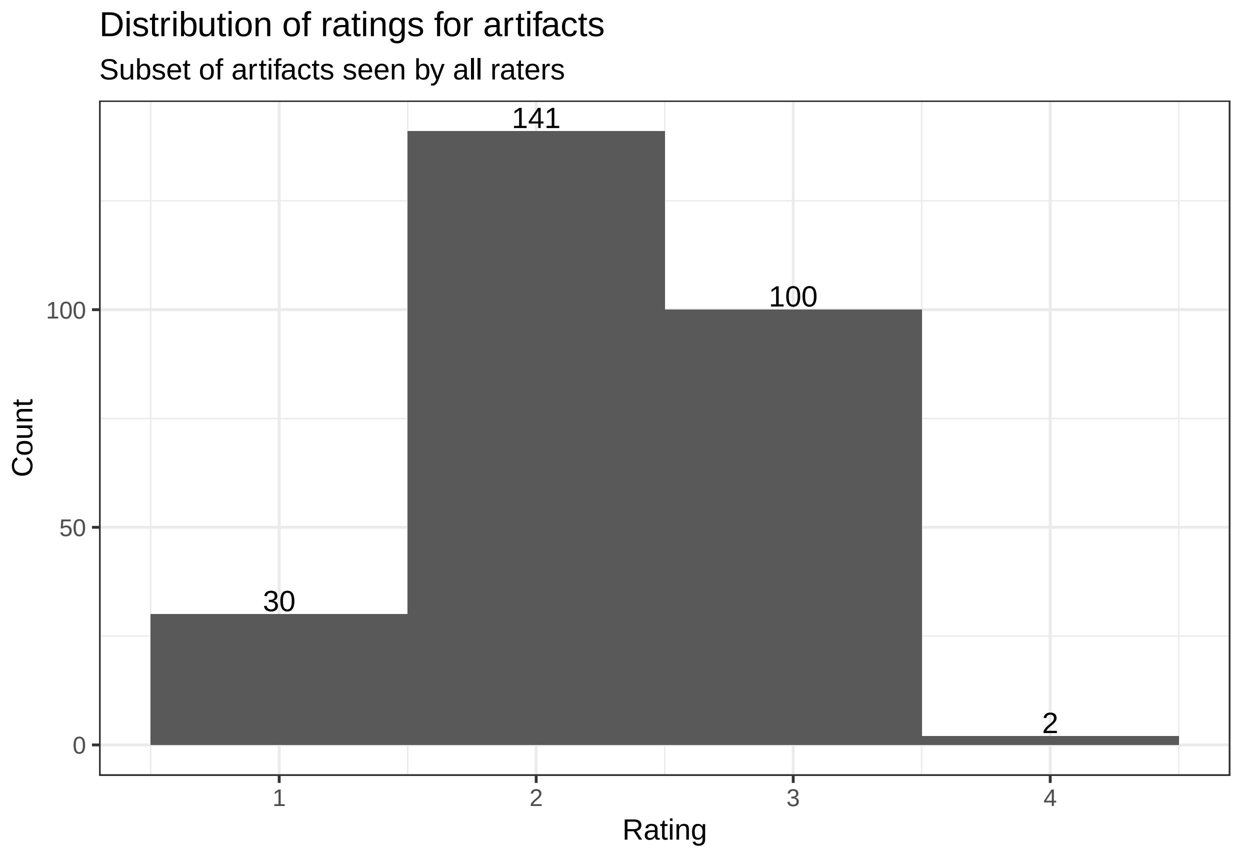
```r
Rater.3 <- tall.full %>%
  mutate(Rater = as.factor(Rater)) %>%
  filter(Rater == "3") %>%
  dplyr::select(c(Rater, Rating))
summary(Rater.3)
```

```
##  Rater       Rating
##  1:  0   Min.   :1.000
##  2:  0   1st Qu.:2.000
##  3:273   Median :2.000
##           Mean   :2.176
##           3rd Qu.:3.000
##           Max.   :4.000
```

```r
sd(Rater.3$Rating)
```

```
## [1] 0.6901631
```

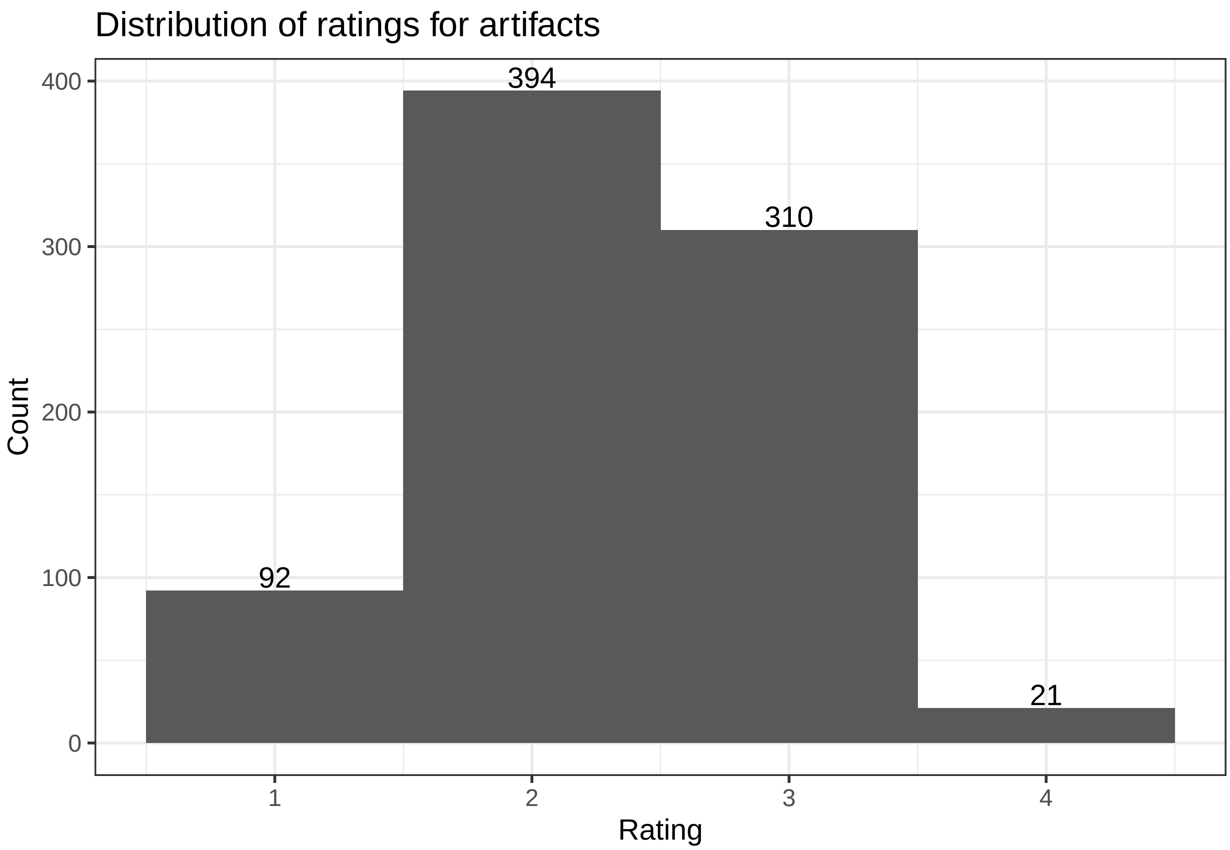
```r
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",
       subtitle = "Subset of artifacts seen by all raters") +
  theme_bw()
```



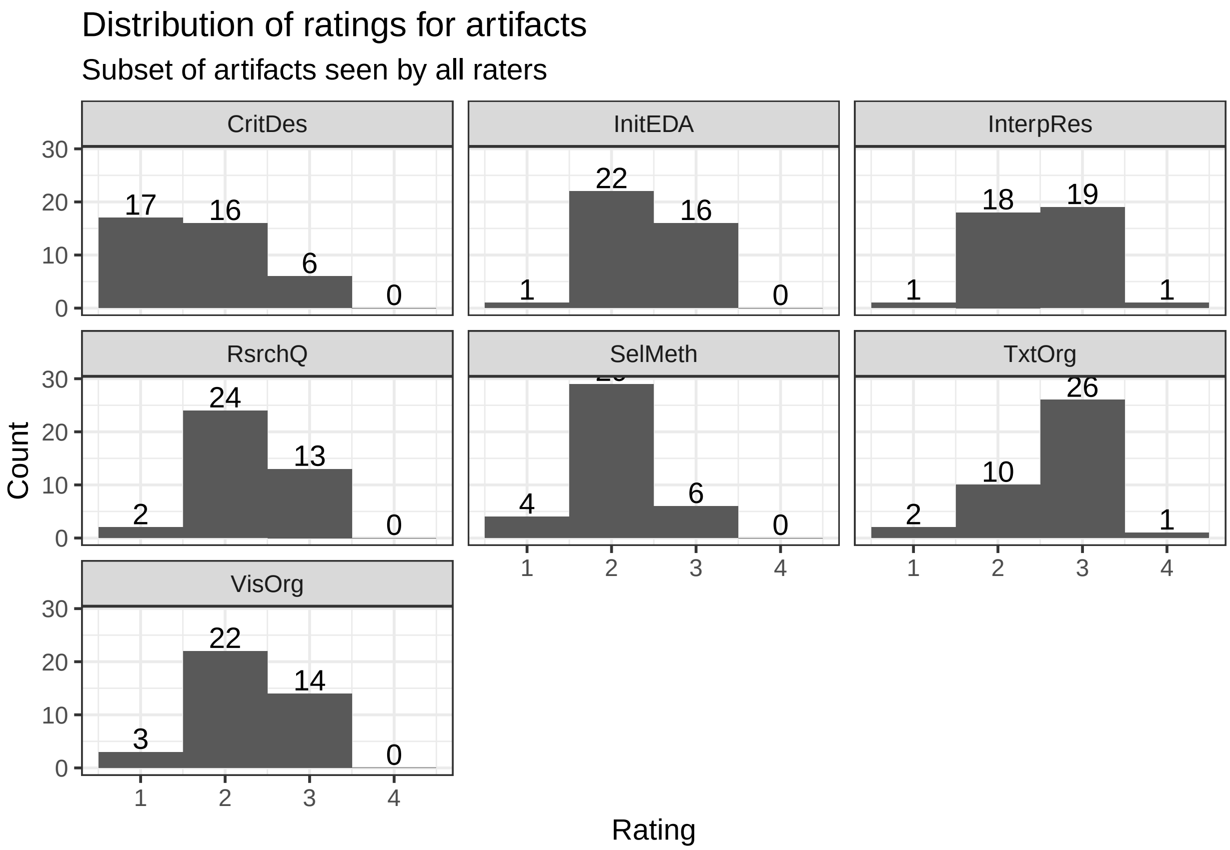

```
tall.full %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw()
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).

## Warning: Removed 2 rows containing non-finite values (stat_bin).
```



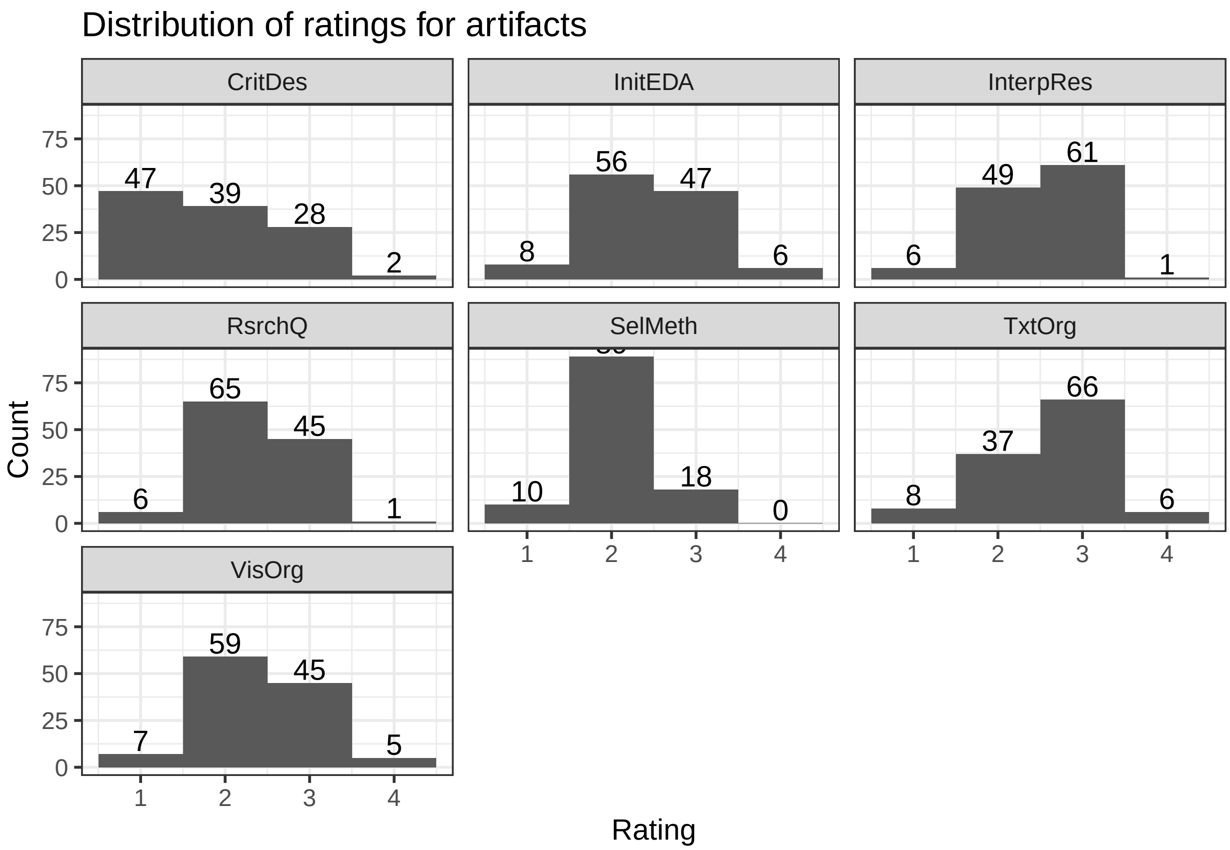

```
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",
       subtitle = "Subset of artifacts seen by all raters") +
  theme_bw() +
  facet_wrap(~ Rubric)
```



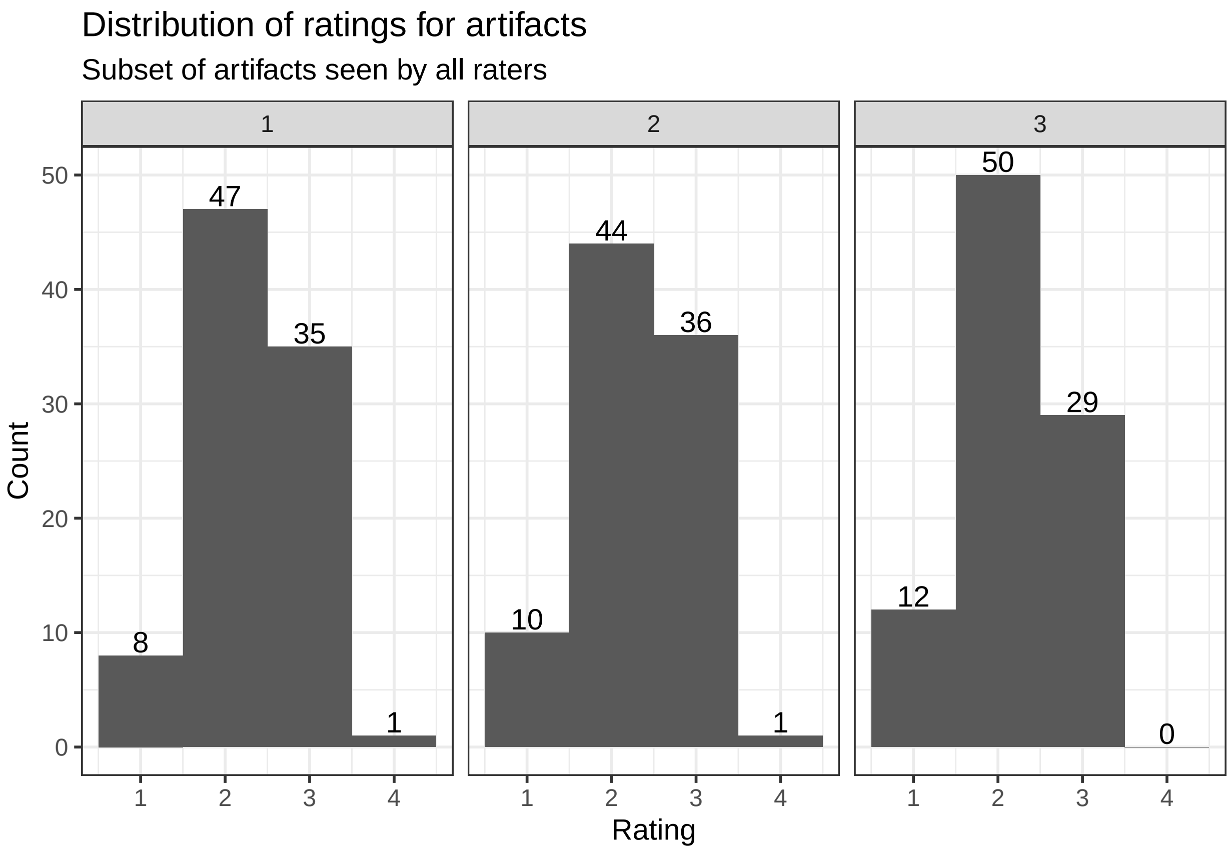

```
tall.full %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw() +
  facet_wrap(~ Rubric)
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).

## Warning: Removed 2 rows containing non-finite values (stat_bin).
```



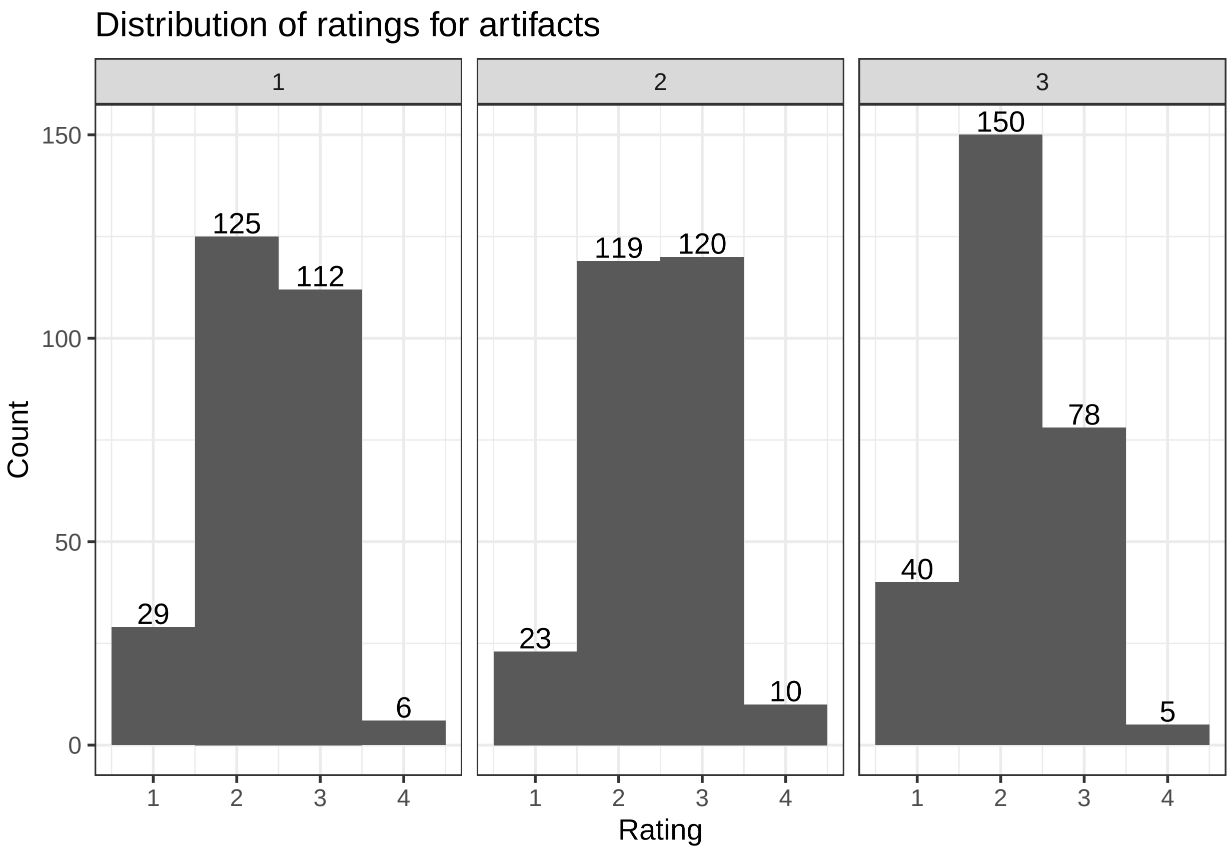

```r
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",
       subtitle = "Subset of artifacts seen by all raters") +
  theme_bw() +
  facet_wrap(~ Rater)
```




```
tall.full %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw() +
  facet_wrap(~ Rater)
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).

## Warning: Removed 2 rows containing non-finite values (stat_bin).
```



We performed EDA to compare the distribution of ratings overall, the distribution of ratings for each of the rubrics, and the distribution of ratings among the raters. This was done for both the subset of data and the full data set.

Appendix 2. Intraclass Correlation and Percent Exact Agreement

Intraclass Correlation per Rubric (Subset Data)

After EDA we wanted to measure the intraclass correlation between raters (ICC). The ICC's help us determine whether the raters are generally in agreement or not on each rubric. To do this we used the subset data and modified it more to only include the specific rubric of interest. We then fit seven random-intercept models, one for each rubric, and then manually calculated the seven ICC's.

```
RsrchQ.ratings <- tall[tall$Rubric=="RsrchQ",]
RsrchQ.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=RsrchQ.ratings)
summary(RsrchQ.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings
##
## REML criterion at convergence: 66.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.3025 -0.5987 -0.3276  0.9696  1.6472
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.05983  0.2446
##  Residual             0.25641  0.5064
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.2821     0.1057   21.59
```

```
ICC.RsrchQ = (0.05983)/(0.05983+0.25641)
ICC.RsrchQ
```

```
## [1] 0.1891918
```

```
CritDes.ratings <- tall[tall$Rubric=="CritDes",]
CritDes.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=CritDes.ratings)
summary(CritDes.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: CritDes.ratings
##
## REML criterion at convergence: 75.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9647 -0.4386 -0.2978  0.5318  2.1987
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3091   0.5560
##  Residual             0.2308   0.4804
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   1.7179     0.1723   9.969
```

```
ICC.CritDes = (0.3091)/(0.3091+0.2308)
ICC.CritDes
```

```
## [1] 0.5725134
```

```
InitEDA.ratings <- tall[tall$Rubric=="InitEDA",]
InitEDA.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=InitEDA.ratings)
summary(InitEDA.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InitEDA.ratings
##
## REML criterion at convergence: 56.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1670 -0.2504 -0.2504  0.4006  1.6663
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1496   0.3867
##  Residual             0.1538   0.3922
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.3846     0.1243   19.18
```

```r
ICC.InitEDA = (0.1496)/(0.1496+ 0.1538)
ICC.InitEDA
```

```
## [1] 0.4930784
```

```r
SelMeth.ratings <- tall[tall$Rubric=="SelMeth",]
SelMeth.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=SelMeth.ratings)
summary(SelMeth.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: SelMeth.ratings
##
## REML criterion at convergence: 50.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.11366 -0.03357 -0.03357  0.62101  2.04652
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1396   0.3736
##  Residual             0.1282   0.3581
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.0513     0.1184   17.32
```

```r
ICC.SelMeth = (0.1396)/(0.1396+0.1396)
ICC.SelMeth
```

```
## [1] 0.5
```

```r
InterpRes.ratings <- tall[tall$Rubric=="InterpRes",]
InterpRes.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=InterpRes.ratings)
summary(InterpRes.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings
##
## REML criterion at convergence: 71.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0965 -0.8061  0.4844  0.7806  2.6635
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08405  0.2899
##  Residual             0.28205  0.5311
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    2.513      0.117   21.47
```

```r
ICC.InterpRes = (0.08405)/(0.08405+0.28205)
ICC.InterpRes
```

```
## [1] 0.2295821
```

```r
VisOrg.ratings <- tall[tall$Rubric=="VisOrg",]
VisOrg.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=VisOrg.ratings)
summary(VisOrg.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: VisOrg.ratings
##
## REML criterion at convergence: 60.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5168 -0.7176 -0.1341  0.3414  1.7241
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2236   0.4729
##  Residual             0.1538   0.3922
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.2821     0.1454   15.69
```

```
ICC.VisOrg = (0.2236)/(0.2236+0.1538)
ICC.VisOrg
```

```
## [1] 0.5924748
```

```
TxtOrg.ratings <- tall[tall$Rubric=="TxtOrg",]
TxtOrg.lmer <- lmer(Rating ~ 1 + (1|Artifact), data=TxtOrg.ratings)
summary(TxtOrg.lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings
##
## REML criterion at convergence: 74.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6943 -0.7698  0.3849  0.3849  2.5019
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.05556  0.2357
##  Residual             0.33333  0.5774
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.6667     0.1132   23.55
```

```
ICC.TxtOrg = (0.05556)/(0.05556+0.33333)
ICC.TxtOrg
```

```
## [1] 0.1428682
```

While ICC's are an important factor and can tell us about general agreement between raters, it cannot tell us which raters might be contributing to disagreement.

Percent Exact Agreement

This leads to the next portion where we calculated the percent exact agreement between combinations of raters. This will help to determine who is agreeing with whom on each rubric. To do this we made 2-way tables of counts for the ratings for each pair of raters on each rubric and divided the counts on the main diagonal by the total number of counts.

```
repeated <- ratings[ratings$Repeated==1,]

RsrchQ.raters12 <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1],
                              r2=repeated$RsrchQ[repeated$Rater==2],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a2=repeated$Artifact[repeated$Rater==2])
```

```r
#with(RsrchQ.raters12,table(r1,r2))
r1 <- factor(RsrchQ.raters12$r1,levels=1:4)
r2 <- factor(RsrchQ.raters12$r2,levels=1:4)
(RQ.12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 0 0 0 0
##   2 1 4 3 0
##   3 1 3 1 0
##   4 0 0 0 0
```

```r
sum(diag(RQ.12)/sum(RQ.12))
```

```
## [1] 0.3846154
```

```r
CritDes.raters12 <- data.frame(r1=repeated$CritDes[repeated$Rater==1],
                               r2=repeated$CritDes[repeated$Rater==2],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a2=repeated$Artifact[repeated$Rater==2])
#with(CritDes.raters12,table(r1,r2))
r1 <- factor(CritDes.raters12$r1,levels=1:4)
r2 <- factor(CritDes.raters12$r2,levels=1:4)
(CD.12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 3 2 1 0
##   2 2 3 1 0
##   3 0 0 1 0
##   4 0 0 0 0
```

```r
sum(diag(CD.12)/sum(CD.12))
```

```
## [1] 0.5384615
```

```r
InitEDA.raters12 <- data.frame(r1=repeated$InitEDA[repeated$Rater==1],
                               r2=repeated$InitEDA[repeated$Rater==2],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a2=repeated$Artifact[repeated$Rater==2])
#with(InitEDA.raters12,table(r1,r2))
r1 <- factor(InitEDA.raters12$r1,levels=1:4)
r2 <- factor(InitEDA.raters12$r2,levels=1:4)
(IE.12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 0 1 0 0
##   2 0 4 0 0
##   3 0 3 5 0
##   4 0 0 0 0
```

```r
sum(diag(IE.12)/sum(IE.12))
```

```
## [1] 0.6923077
```

```r
SelMeth.raters12 <- data.frame(r1=repeated$SelMeth[repeated$Rater==1],
                               r2=repeated$SelMeth[repeated$Rater==2],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a2=repeated$Artifact[repeated$Rater==2])
#with(SelMeth.raters12,table(r1,r2))
r1 <- factor(SelMeth.raters12$r1,levels=1:4)
r2 <- factor(SelMeth.raters12$r2,levels=1:4)
(SM.12 <- table(r1,r2))
```

```
##     r2
## r1    1  2  3  4
##    1  0  0  0  0
##    2  1 10  0  0
##    3  0  0  2  0
##    4  0  0  0  0
```

```r
sum(diag(SM.12)/sum(SM.12))
```

```
## [1] 0.9230769
```

```r
InterpRes.raters12 <- data.frame(r1=repeated$InterpRes[repeated$Rater==1],
                                 r2=repeated$InterpRes[repeated$Rater==2],
                                 a1=repeated$Artifact[repeated$Rater==1],
                                 a2=repeated$Artifact[repeated$Rater==2])
#with(InterpRes.raters12,table(r1,r2))
r1 <- factor(InterpRes.raters12$r1,levels=1:4)
r2 <- factor(InterpRes.raters12$r2,levels=1:4)
(IR.12 <- table(r1,r2))
```

```
##     r2
## r1   1 2 3 4
##    1 0 0 0 0
##    2 0 3 1 1
##    3 0 3 5 0
##    4 0 0 0 0
```

```r
sum(diag(IR.12)/sum(IR.12))
```

```
## [1] 0.6153846
```

```r
VisOrg.raters12 <- data.frame(r1=repeated$VisOrg[repeated$Rater==1],
                              r2=repeated$VisOrg[repeated$Rater==2],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a2=repeated$Artifact[repeated$Rater==2])
#with(VisOrg.raters12,table(r1,r2))
r1 <- factor(VisOrg.raters12$r1,levels=1:4)
r2 <- factor(VisOrg.raters12$r2,levels=1:4)
(VO.12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 1 0 0 0
##   2 0 4 5 0
##   3 0 1 2 0
##   4 0 0 0 0
```

```
sum(diag(VO.12)/sum(VO.12))
```

```
## [1] 0.5384615
```

```
TxtOrg.raters12 <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
                              r2=repeated$TxtOrg[repeated$Rater==2],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a2=repeated$Artifact[repeated$Rater==2])
#with(TxtOrg.raters12,table(r1,r2))
r1 <- factor(TxtOrg.raters12$r1,levels=1:4)
r2 <- factor(TxtOrg.raters12$r2,levels=1:4)
(TO.12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 0 0 0 0
##   2 0 2 2 0
##   3 0 1 7 0
##   4 1 0 0 0
```

```
sum(diag(TO.12)/sum(TO.12))
```

```
## [1] 0.6923077
```

```
RsrchQ.raters13 <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1],
                              r3=repeated$RsrchQ[repeated$Rater==3],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(RsrchQ.raters13,table(r1,r3))
r1 <- factor(RsrchQ.raters13$r1,levels=1:4)
r3 <- factor(RsrchQ.raters13$r3,levels=1:4)
(RQ.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 0 0 0
##   2 0 7 1 0
##   3 0 2 3 0
##   4 0 0 0 0
```

```
sum(diag(RQ.13)/sum(RQ.13))
```

```
## [1] 0.7692308
```

```
CritDes.raters13 <- data.frame(r1=repeated$CritDes[repeated$Rater==1],
                               r3=repeated$CritDes[repeated$Rater==3],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(CritDes.raters13,table(r1,r3))
r1 <- factor(CritDes.raters13$r1,levels=1:4)
r3 <- factor(CritDes.raters13$r3,levels=1:4)
(CD.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 4 2 0 0
##   2 2 3 1 0
##   3 0 0 1 0
##   4 0 0 0 0
```

```
sum(diag(CD.13)/sum(CD.13))
```

```
## [1] 0.6153846
```

```
InitEDA.raters13 <- data.frame(r1=repeated$InitEDA[repeated$Rater==1],
                               r3=repeated$InitEDA[repeated$Rater==3],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(InitEDA.raters13,table(r1,r3))
r1 <- factor(InitEDA.raters13$r1,levels=1:4)
r3 <- factor(InitEDA.raters13$r3,levels=1:4)
(IE.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 1 0 0
##   2 0 4 0 0
##   3 0 5 3 0
##   4 0 0 0 0
```

```
sum(diag(IE.13)/sum(IE.13))
```

```
## [1] 0.5384615
```

```
SelMeth.raters13 <- data.frame(r1=repeated$SelMeth[repeated$Rater==1],
                               r3=repeated$SelMeth[repeated$Rater==3],
                               a1=repeated$Artifact[repeated$Rater==1],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(SelMeth.raters13,table(r1,r3))
r1 <- factor(SelMeth.raters13$r1,levels=1:4)
r3 <- factor(SelMeth.raters13$r3,levels=1:4)
(SM.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 0 0 0
##   2 3 7 1 0
##   3 0 1 1 0
##   4 0 0 0 0
```

```
sum(diag(SM.13)/sum(SM.13))
```

```
## [1] 0.6153846
```

```
InterpRes.raters13 <- data.frame(r1=repeated$InterpRes[repeated$Rater==1],
                                 r3=repeated$InterpRes[repeated$Rater==3],
                                 a1=repeated$Artifact[repeated$Rater==1],
                                 a3=repeated$Artifact[repeated$Rater==3])
#with(InterpRes.raters13,table(r1,r3))
r1 <- factor(InterpRes.raters13$r1,levels=1:4)
r3 <- factor(InterpRes.raters13$r3,levels=1:4)
(IR.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 0 0 0
##   2 1 3 1 0
##   3 0 4 4 0
##   4 0 0 0 0
```

```
sum(diag(IR.13)/sum(IR.13))
```

```
## [1] 0.5384615
```

```
VisOrg.raters13 <- data.frame(r1=repeated$VisOrg[repeated$Rater==1],
                              r3=repeated$VisOrg[repeated$Rater==3],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(VisOrg.raters13,table(r1,r3))
r1 <- factor(VisOrg.raters13$r1,levels=1:4)
r3 <- factor(VisOrg.raters13$r3,levels=1:4)
(VO.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 1 0 0 0
##   2 0 7 2 0
##   3 0 1 2 0
##   4 0 0 0 0
```

```
sum(diag(VO.13)/sum(VO.13))
```

```
## [1] 0.7692308
```

```
TxtOrg.raters13 <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
                              r3=repeated$TxtOrg[repeated$Rater==3],
                              a1=repeated$Artifact[repeated$Rater==1],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(TxtOrg.raters13,table(r1,r3))
r1 <- factor(TxtOrg.raters13$r1,levels=1:4)
r3 <- factor(TxtOrg.raters13$r3,levels=1:4)
(TO.13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 0 0 0
##   2 1 1 2 0
##   3 0 1 7 0
##   4 0 1 0 0
```

```
sum(diag(TO.13)/sum(TO.13))
```

```
## [1] 0.6153846
```

```
RsrchQ.raters23 <- data.frame(r2=repeated$RsrchQ[repeated$Rater==2],
                              r3=repeated$RsrchQ[repeated$Rater==3],
                              a2=repeated$Artifact[repeated$Rater==2],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(RsrchQ.raters23,table(r2,r3))
r2 <- factor(RsrchQ.raters23$r2,levels=1:4)
r3 <- factor(RsrchQ.raters23$r3,levels=1:4)
(RQ.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 2 0 0
##   2 0 5 2 0
##   3 0 2 2 0
##   4 0 0 0 0
```

```
sum(diag(RQ.23)/sum(RQ.23))
```

```
## [1] 0.5384615
```

```
CritDes.raters23 <- data.frame(r2=repeated$CritDes[repeated$Rater==2],
                               r3=repeated$CritDes[repeated$Rater==3],
                               a2=repeated$Artifact[repeated$Rater==2],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(CritDes.raters23,table(r2,r3))
r2 <- factor(CritDes.raters23$r2,levels=1:4)
r3 <- factor(CritDes.raters23$r3,levels=1:4)
(CD.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 5 0 0 0
##   2 1 3 1 0
##   3 0 2 1 0
##   4 0 0 0 0
```

```
sum(diag(RQ.23)/sum(RQ.23))
```

```
## [1] 0.5384615
```

```
InitEDA.raters23 <- data.frame(r2=repeated$InitEDA[repeated$Rater==2],
                               r3=repeated$InitEDA[repeated$Rater==3],
                               a2=repeated$Artifact[repeated$Rater==2],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(InitEDA.raters23,table(r2,r3))
r2 <- factor(InitEDA.raters23$r2,levels=1:4)
r3 <- factor(InitEDA.raters23$r3,levels=1:4)
(IE.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 0 0 0
##   2 0 8 0 0
##   3 0 2 3 0
##   4 0 0 0 0
```

```
sum(diag(IE.23)/sum(IE.23))
```

```
## [1] 0.8461538
```

```
SelMeth.raters23 <- data.frame(r2=repeated$SelMeth[repeated$Rater==2],
                               r3=repeated$SelMeth[repeated$Rater==3],
                               a2=repeated$Artifact[repeated$Rater==2],
                               a3=repeated$Artifact[repeated$Rater==3])
#with(SelMeth.raters23,table(r2,r3))
r2 <- factor(SelMeth.raters23$r2,levels=1:4)
r3 <- factor(SelMeth.raters23$r3,levels=1:4)
(SM.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 1 0 0 0
##   2 2 7 1 0
##   3 0 1 1 0
##   4 0 0 0 0
```

```
sum(diag(SM.23)/sum(SM.23))
```

```
## [1] 0.6923077
```

```r
InterpRes.raters23 <- data.frame(r2=repeated$InterpRes[repeated$Rater==2],
                                 r3=repeated$InterpRes[repeated$Rater==3],
                                 a2=repeated$Artifact[repeated$Rater==2],
                                 a3=repeated$Artifact[repeated$Rater==3])
#with(InterpRes.raters23,table(r2,r3))
r2 <- factor(InterpRes.raters23$r2,levels=1:4)
r3 <- factor(InterpRes.raters23$r3,levels=1:4)
(IR.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 0 0 0
##   2 1 4 1 0
##   3 0 2 4 0
##   4 0 1 0 0
```

```r
sum(diag(IR.23)/sum(IR.23))
```

```
## [1] 0.6153846
```

```r
VisOrg.raters23 <- data.frame(r2=repeated$VisOrg[repeated$Rater==2],
                              r3=repeated$VisOrg[repeated$Rater==3],
                              a2=repeated$Artifact[repeated$Rater==2],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(VisOrg.raters23,table(r2,r3))
r2 <- factor(VisOrg.raters23$r2,levels=1:4)
r3 <- factor(VisOrg.raters23$r3,levels=1:4)
(VO.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 1 0 0 0
##   2 0 5 0 0
##   3 0 3 4 0
##   4 0 0 0 0
```

```r
sum(diag(VO.23)/sum(VO.23))
```

```
## [1] 0.7692308
```

```r
TxtOrg.raters23 <- data.frame(r2=repeated$TxtOrg[repeated$Rater==2],
                              r3=repeated$TxtOrg[repeated$Rater==3],
                              a2=repeated$Artifact[repeated$Rater==2],
                              a3=repeated$Artifact[repeated$Rater==3])
#with(TxtOrg.raters23,table(r2,r3))
r2 <- factor(TxtOrg.raters23$r2,levels=1:4)
r3 <- factor(TxtOrg.raters23$r3,levels=1:4)
(TO.23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 1 0 0
##   2 1 0 2 0
##   3 0 2 7 0
##   4 0 0 0 0
```

```
sum(diag(TO.23)/sum(TO.23))
```

```
## [1] 0.5384615
```

Intraclass Correlation per Rubric (Full Data)

We repeated the process of calculating ICC's for the full data set. We are not able to do the same for the percent exact agreements because not all raters rated every artifact.

```
RsrchQ.ratings.full <- tall.full[tall.full$Rubric=="RsrchQ",]
RsrchQ.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=RsrchQ.ratings.full)
summary(RsrchQ.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings.full
##
## REML criterion at convergence: 211.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2748 -0.5365 -0.3780  0.9626  2.4617
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07372  0.2715
##  Residual             0.27797  0.5272
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.35790    0.05774   40.84
```

```
ICC.RsrchQ.full = (0.07372)/(0.07372+0.27797)
ICC.RsrchQ.full
```

```
## [1] 0.2096164
```

```
CritDes.ratings.full <- tall.full[tall.full$Rubric=="CritDes",]
CritDes.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=CritDes.ratings.full)
summary(CritDes.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: CritDes.ratings.full
##
## REML criterion at convergence: 277.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.01042 -0.60409  0.04407  0.72769  2.06310
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4963   0.7045
##  Residual             0.2411   0.4910
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.90720    0.08874   21.49
```

```r
ICC.CritDes.full = (0.4963)/(0.4963+0.2411)
ICC.CritDes.full
```

```
## [1] 0.6730404
```

```r
InitEDA.ratings.full <- tall.full[tall.full$Rubric=="InitEDA",]
InitEDA.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=InitEDA.ratings.full)
summary(InitEDA.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InitEDA.ratings.full
##
## REML criterion at convergence: 240.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8923 -0.3451 -0.1454  0.4250  1.6015
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3628   0.6023
##  Residual             0.1655   0.4068
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44815    0.07479   32.73
```

```r
ICC.InitEDA.full = (0.3628)/(0.3628+0.1655)
ICC.InitEDA.full
```

```
## [1] 0.686731
```

```r
SelMeth.ratings.full <- tall.full[tall.full$Rubric=="SelMeth",]
SelMeth.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=SelMeth.ratings.full)
summary(SelMeth.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: SelMeth.ratings.full
##
## REML criterion at convergence: 157.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2057 -0.1075 -0.1075 -0.0553  2.0951
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1108   0.3329
##  Residual             0.1240   0.3521
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.07168    0.04893   42.34
```

```r
ICC.SelMeth.full = (0.1108)/(0.1108+0.1240)
ICC.SelMeth.full
```

```
## [1] 0.471891
```

```r
InterpRes.ratings.full <- tall.full[tall.full$Rubric=="InterpRes",]
InterpRes.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=InterpRes.ratings.full)
summary(InterpRes.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings.full
##
## REML criterion at convergence: 217.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1448 -0.6998  0.5175  0.7452  2.6532
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08219  0.2867
##  Residual             0.29136  0.5398
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.48427    0.05962   41.67
```

```r
ICC.InterpRes.full = (0.08219)/(0.08219+0.29136)
ICC.InterpRes.full
```

```
## [1] 0.2200241
```

```r
VisOrg.ratings.full <- tall.full[tall.full$Rubric=="VisOrg",]
VisOrg.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=VisOrg.ratings.full)
summary(VisOrg.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: VisOrg.ratings.full
##
## REML criterion at convergence: 226.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5918 -0.3789 -0.1632  0.4726  1.6322
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3092   0.5561
##  Residual             0.1588   0.3985
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44497    0.07063   34.62
```

```r
ICC.VisOrg.full = (0.3092)/(0.3092+0.1588)
ICC.VisOrg.full
```

```
## [1] 0.6606838
```

```r
TxtOrg.ratings.full <- tall.full[tall.full$Rubric=="TxtOrg",]
TxtOrg.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=TxtOrg.ratings.full)
summary(TxtOrg.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings.full
##
## REML criterion at convergence: 249
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3638 -0.7641  0.3836  0.5278  2.4094
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09145  0.3024
```

```
##  Residual             0.39503  0.6285
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.59144    0.06764   38.31
```

```
ICC.TxtOrg.full = (0.09145)/(0.09145+0.39503)
ICC.TxtOrg.full
```

```
## [1] 0.1879831
```

The ICC's for each rubric for the full data set are pretty similar to the ICC's for the subset of data. They are a little higher in some cases and I believe a bit more accurate since there are more observations.

Appendix 3. Variable Selection

Fixed Effects per Rubric (Full Data)

Now we want to find out how other variables in the data set interact with ratings. We performed manual variable selection models on each of the seven rubrics for the full data set.

For fixed effect models we used AIC/BIC and likelihood ration tests to compare models to find the best one for each rubric. We also calculated the ICC's again.

```
tall.full.nonmissing <- tall.full[-c(161,684),]
tall.full.nonmissing <- tall.full.nonmissing[tall.full.nonmissing$Sex!="",]


RsrchQ.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="RsrchQ",]
RsrchQ.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=RsrchQ.ratings.full1)
summary(RsrchQ.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings.full1
##
## REML criterion at convergence: 211.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2748 -0.5365 -0.3780  0.9626  2.4617
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07372  0.2715
##  Residual             0.27797  0.5272
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.35790    0.05774   40.84
```

```r
RsrchQ.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex +
                             (1|Artifact), data=RsrchQ.ratings.full1)
summary(RsrchQ.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: RsrchQ.ratings.full1
##
## REML criterion at convergence: 215.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3050 -0.5381 -0.2972  0.8643  2.4434
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06873  0.2622
##  Residual             0.28470  0.5336
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  3.18469    0.60865   5.232
## as.factor(Rater)2  3.10114    0.60862   5.095
## as.factor(Rater)3  3.00000    0.59450   5.046
## SemesterS19        0.08684    0.13225   0.657
## SexF              -0.74473    0.60721  -1.226
## SexM              -0.80317    0.60567  -1.326
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.977
## as.fctr(R)3  0.977  0.977
## SemesterS19  0.000  0.002  0.000
## SexF        -0.981 -0.980 -0.979 -0.092
## SexM        -0.981 -0.982 -0.982 -0.034  0.980
```

```r
anova(RsrchQ.lmer.full,RsrchQ.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: RsrchQ.ratings.full1
## Models:
## RsrchQ.lmer.full: Rating ~ 1 + (1 | Artifact)
## RsrchQ.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## RsrchQ.lmer.full        3 213.19 221.48 -103.60   207.19
## RsrchQ.lmer.allfixed    8 218.82 240.92 -101.41   202.82 4.3689  5     0.4976
```

```r
RsrchQ.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                           data=RsrchQ.ratings.full1)
anova(RsrchQ.lmer.full,RsrchQ.lmer.fixed.3)
```

	df	AIC	BIC
RsrchQ.lmer.full	3	217.0659	225.3524
RsrchQ.lmer.allfixed	8	231.4872	253.5846
RsrchQ.lmer.fixed.3	6	227.8328	244.4058
RsrchQ.lmer.fixed.2	5	224.0995	237.9103

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: RsrchQ.ratings.full1
## Models:
## RsrchQ.lmer.full: Rating ~ 1 + (1 | Artifact)
## RsrchQ.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                     npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## RsrchQ.lmer.full       3 213.19 221.48 -103.6   207.19
## RsrchQ.lmer.fixed.3    6 216.81 233.38 -102.4   204.81 2.3856  3     0.4963
```

```
RsrchQ.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                           data=RsrchQ.ratings.full1)
anova(RsrchQ.lmer.full,RsrchQ.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: RsrchQ.ratings.full1
## Models:
## RsrchQ.lmer.full: Rating ~ 1 + (1 | Artifact)
## RsrchQ.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##                     npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## RsrchQ.lmer.full       3 213.19 221.48 -103.6   207.19
## RsrchQ.lmer.fixed.2    5 215.39 229.20 -102.7   205.39 1.8013  2     0.4063
```

```
ICC.RsrchQ.fixed = ( 0.07276)/( 0.07276+0.27825)
ICC.RsrchQ.fixed
```

```
## [1] 0.2072875
```

```
data.frame(AIC=AIC(RsrchQ.lmer.full,RsrchQ.lmer.allfixed,RsrchQ.lmer.fixed.3,RsrchQ.lmer.fixed.2),
           BIC(RsrchQ.lmer.full,RsrchQ.lmer.allfixed,RsrchQ.lmer.fixed.3,RsrchQ.lmer.fixed.2))[,-3] %>%
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```
CritDes.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="CritDes",]
CritDes.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=CritDes.ratings.full1)
summary(CritDes.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: CritDes.ratings.full1
##
## REML criterion at convergence: 277.9
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.01042 -0.60409  0.04407  0.72769  2.06310
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4963   0.7045
##  Residual             0.2411   0.4910
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.90720    0.08874   21.49
```

```
CritDes.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                             data=CritDes.ratings.full1)
summary(CritDes.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: CritDes.ratings.full1
##
## REML criterion at convergence: 273.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.58234 -0.54151 -0.02948  0.59834  1.59203
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4462   0.6680
##  Residual             0.2477   0.4977
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.80090    0.84639   3.309
## as.factor(Rater)2  3.22089    0.84648   3.805
## as.factor(Rater)3  3.00000    0.83302   3.601
## SemesterS19       -0.03638    0.19723  -0.184
## SexF              -1.14341    0.85004  -1.345
## SexM              -1.04596    0.84771  -1.234
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.984
## as.fctr(R)3  0.984  0.984
## SemesterS19  0.000  0.004  0.000
## SexF        -0.981 -0.980 -0.980 -0.099
## SexM        -0.982 -0.983 -0.983 -0.036  0.977
```

```r
anova(CritDes.lmer.full,CritDes.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings.full1
## Models:
## CritDes.lmer.full: Rating ~ 1 + (1 | Artifact)
## CritDes.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes.lmer.full        3 280.86 289.12 -137.43   274.86
## CritDes.lmer.allfixed    8 280.62 302.64 -132.31   264.62 10.242  5    0.06867
##
## CritDes.lmer.full
## CritDes.lmer.allfixed .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
CritDes.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                            data=CritDes.ratings.full1)
summary(CritDes.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: CritDes.ratings.full1
##
## REML criterion at convergence: 275.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.55736 -0.51636 -0.04172  0.61602  1.61058
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4456   0.6675
##  Residual             0.2479   0.4979
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  1.71921    0.13474  12.760
## as.factor(Rater)2  2.14197    0.13369  16.021
## as.factor(Rater)3  1.93910    0.13345  14.530
## SemesterS19       -0.08386    0.18802  -0.446
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.377
## as.fctr(R)3  0.381  0.372
## SemesterS19 -0.435 -0.399 -0.416
```

```
anova(CritDes.lmer.full,CritDes.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: CritDes.ratings.full1
## Models:
## CritDes.lmer.full: Rating ~ 1 + (1 | Artifact)
## CritDes.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes.lmer.full       3 280.86 289.12 -137.43   274.86
## CritDes.lmer.fixed.3    6 278.65 295.18 -133.33   266.65 8.2035  3    0.04199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
CritDes.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                            data=CritDes.ratings.full1)
summary(CritDes.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: CritDes.ratings.full1
##
## REML criterion at convergence: 274.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.54697 -0.50107 -0.08068  0.63782  1.61697
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4401   0.6634
##  Residual             0.2475   0.4975
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   1.6926     0.1210   13.99
## as.factor(Rater)2   2.1184     0.1222   17.34
## as.factor(Rater)3   1.9144     0.1210   15.83
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.245
## as.fctr(R)3 0.243  0.245
```

```
anova(CritDes.lmer.full,CritDes.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: CritDes.ratings.full1
## Models:
```

	df	AIC	BIC
CritDes.lmer.full	3	283.8691	292.1299
CritDes.lmer.allfixed	8	289.6376	311.6663
CritDes.lmer.fixed.3	6	287.5167	304.0383
CritDes.lmer.fixed.2	5	284.2078	297.9758

```
## CritDes.lmer.full: Rating ~ 1 + (1 | Artifact)
## CritDes.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##                        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes.lmer.full       3 280.86 289.12 -137.43   274.86
## CritDes.lmer.fixed.2    5 276.86 290.62 -133.43   266.86 7.9996  2    0.01832 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ICC.CritDes.fixed = (0.4909)/(0.4909+0.2412)
ICC.CritDes.fixed
```

```
## [1] 0.6705368
```

```
data.frame(AIC=AIC(CritDes.lmer.full,CritDes.lmer.allfixed,CritDes.lmer.fixed.3,CritDes.lmer.fixed.2),
           BIC(CritDes.lmer.full,CritDes.lmer.allfixed,CritDes.lmer.fixed.3,CritDes.lmer.fixed.2))[,-3]
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```
InitEDA.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="InitEDA",]
InitEDA.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=InitEDA.ratings.full1)
summary(InitEDA.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InitEDA.ratings.full1
##
## REML criterion at convergence: 240.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8923 -0.3451 -0.1454  0.4250  1.6015
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3628   0.6023
##  Residual             0.1655   0.4068
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44815    0.07479   32.73
```

```r
InitEDA.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                              data=InitEDA.ratings.full1)
summary(InitEDA.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: InitEDA.ratings.full1
##
## REML criterion at convergence: 244.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.07349 -0.35275 -0.02661  0.38131  1.51459
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3878   0.6227
##  Residual             0.1548   0.3934
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  3.193322   0.746910   4.275
## as.factor(Rater)2  3.205687   0.746893   4.292
## as.factor(Rater)3  3.000000   0.736598   4.073
## SemesterS19       -0.008809   0.173677  -0.051
## SexF              -0.710386   0.751240  -0.946
## SexM              -0.662853   0.749169  -0.885
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.986
## as.fctr(R)3  0.986  0.986
## SemesterS19  0.000  0.001  0.000
## SexF        -0.981 -0.981 -0.981 -0.101
## SexM        -0.983 -0.983 -0.983 -0.035  0.977
```

```r
anova(InitEDA.lmer.full,InitEDA.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InitEDA.ratings.full1
## Models:
## InitEDA.lmer.full: Rating ~ 1 + (1 | Artifact)
## InitEDA.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA.lmer.full        3 243.42 251.71 -118.71   237.42
## InitEDA.lmer.allfixed    8 249.26 271.36 -116.63   233.26 4.1599  5     0.5266
```

```r
InitEDA.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                             data=InitEDA.ratings.full1)
summary(InitEDA.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: InitEDA.ratings.full1
##
## REML criterion at convergence: 244.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.07179 -0.37862 -0.01407  0.38323  1.51216
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3801   0.6165
##  Residual             0.1555   0.3944
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.51421    0.11573  21.726
## as.factor(Rater)2  2.52944    0.11471  22.050
## as.factor(Rater)3  2.33320    0.11471  20.339
## SemesterS19       -0.03469    0.16386  -0.212
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.430
## as.fctr(R)3  0.430  0.425
## SemesterS19 -0.445 -0.429 -0.429
```

```
anova(InitEDA.lmer.full,InitEDA.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InitEDA.ratings.full1
## Models:
## InitEDA.lmer.full: Rating ~ 1 + (1 | Artifact)
## InitEDA.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA.lmer.full       3 243.42 251.71 -118.71   237.42
## InitEDA.lmer.fixed.3    6 246.24 262.81 -117.12   234.24 3.1871  3     0.3637
```

```
InitEDA.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                            data=InitEDA.ratings.full1)
summary(InitEDA.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: InitEDA.ratings.full1
##
## REML criterion at convergence: 242.7
##
## Scaled residuals:
```

	df	AIC	BIC
InitEDA.lmer.full	3	246.7763	255.0628
InitEDA.lmer.allfixed	8	260.0518	282.1492
InitEDA.lmer.fixed.3	6	256.4146	272.9876
InitEDA.lmer.fixed.2	5	252.6742	266.4851

```
##      Min       1Q   Median       3Q      Max
## -2.06664 -0.37458 -0.00988  0.37103  1.51561
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3737   0.6113
##  Residual             0.1560   0.3949
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.5024     0.1032   24.24
## as.factor(Rater)2   2.5196     0.1032   24.41
## as.factor(Rater)3   2.3230     0.1032   22.50
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.292
## as.fctr(R)3 0.292  0.292
```

```
anova(InitEDA.lmer.full,InitEDA.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InitEDA.ratings.full1
## Models:
## InitEDA.lmer.full: Rating ~ 1 + (1 | Artifact)
## InitEDA.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA.lmer.full       3 243.42 251.71 -118.71   237.42
## InitEDA.lmer.fixed.2    5 244.28 258.09 -117.14   234.28 3.1408  2      0.208
```

```
ICC.InitEDA.fixed = (0.3651)/(0.3651+0.1655)
ICC.InitEDA.fixed
```

```
## [1] 0.688089
```

```
data.frame(AIC=AIC(InitEDA.lmer.full,InitEDA.lmer.allfixed,InitEDA.lmer.fixed.3,InitEDA.lmer.fixed.2),
           BIC(InitEDA.lmer.full,InitEDA.lmer.allfixed,InitEDA.lmer.fixed.3,InitEDA.lmer.fixed.2))[,-3]
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```r
SelMeth.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="SelMeth",]
SelMeth.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=SelMeth.ratings.full1)
summary(SelMeth.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 157.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2057 -0.1075 -0.1075 -0.0553  2.0951
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1108   0.3329
##  Residual             0.1240   0.3521
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.07168    0.04893   42.34
```

```r
SelMeth.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                             data=SelMeth.ratings.full1)
summary(SelMeth.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 144.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.09631 -0.34555 -0.06849  0.33489  2.66067
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09013  0.3002
##  Residual             0.10714  0.3273
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   3.2227     0.4531   7.113
## as.factor(Rater)2   3.1946     0.4530   7.051
## as.factor(Rater)3   3.0000     0.4441   6.755
## SemesterS19        -0.3195     0.1025  -3.119
## SexF               -1.0352     0.4536  -2.282
## SexM               -0.9136     0.4523  -2.020
```

```
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.981
## as.fctr(R)3  0.980  0.980
## SemesterS19  0.000  0.002  0.000
## SexF        -0.980 -0.980 -0.979 -0.097
## SexM        -0.981 -0.982 -0.982 -0.035  0.978
```

```
anova(SelMeth.lmer.full,SelMeth.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings.full1
## Models:
## SelMeth.lmer.full: Rating ~ 1 + (1 | Artifact)
## SelMeth.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth.lmer.full        3 159.53 167.82 -76.768   153.53
## SelMeth.lmer.allfixed    8 144.52 166.62 -64.260   128.52 25.015  5  0.0001384
##
## SelMeth.lmer.full
## SelMeth.lmer.allfixed ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
SelMeth.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                            data=SelMeth.ratings.full1)
summary(SelMeth.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 148.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.07037 -0.37872 -0.09732  0.22451  2.50908
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09505  0.3083
##  Residual             0.11002  0.3317
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.25854    0.07612  29.671
## as.factor(Rater)2  2.23392    0.07532  29.658
## as.factor(Rater)3  2.06017    0.07532  27.351
## SemesterS19       -0.37273    0.09954  -3.744
```

```
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.288
## as.fctr(R)3  0.288  0.280
## SemesterS19 -0.412 -0.390 -0.390
```

```
anova(SelMeth.lmer.allfixed,SelMeth.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: SelMeth.ratings.full1
## Models:
## SelMeth.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
## SelMeth.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth.lmer.fixed.3     6 146.92 163.49 -67.461   134.92
## SelMeth.lmer.allfixed    8 144.52 166.62 -64.260   128.52 6.4004  2    0.04075
##
## SelMeth.lmer.fixed.3
## SelMeth.lmer.allfixed *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
SelMeth.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                            data=SelMeth.ratings.full1)
summary(SelMeth.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 158.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8986 -0.2102 -0.1821  0.3038  2.5063
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1144   0.3382
##  Residual             0.1163   0.3410
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.14216    0.07303   29.33
## as.factor(Rater)2  2.12316    0.07303   29.07
## as.factor(Rater)3  1.95001    0.07303   26.70
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.165
## as.fctr(R)3 0.165  0.165
```

```r
anova(SelMeth.lmer.allfixed,SelMeth.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings.full1
## Models:
## SelMeth.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
## SelMeth.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth.lmer.fixed.2     5 158.52 172.33 -74.261   148.52
## SelMeth.lmer.allfixed    8 144.52 166.62 -64.260   128.52 20.002  3  0.0001696
##
## SelMeth.lmer.fixed.2
## SelMeth.lmer.allfixed ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(SelMeth.lmer.full,SelMeth.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings.full1
## Models:
## SelMeth.lmer.full: Rating ~ 1 + (1 | Artifact)
## SelMeth.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth.lmer.full       3 159.53 167.82 -76.768   153.53
## SelMeth.lmer.fixed.3    6 146.92 163.49 -67.461   134.92 18.614  3  0.0003285
##
## SelMeth.lmer.full
## SelMeth.lmer.fixed.3 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ICC.SelMeth.fixed = (0.08973)/(0.08973+0.10842)
ICC.SelMeth.fixed
```

```
## [1] 0.4528388
```

```r
data.frame(AIC=AIC(SelMeth.lmer.full,SelMeth.lmer.allfixed,SelMeth.lmer.fixed.3,SelMeth.lmer.fixed.2),
           BIC(SelMeth.lmer.full,SelMeth.lmer.allfixed,SelMeth.lmer.fixed.3,SelMeth.lmer.fixed.2))[,-3]
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```r
InterpRes.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="InterpRes",]
InterpRes.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=InterpRes.ratings.full1)
summary(InterpRes.lmer.full)
```

	df	AIC	BIC
SelMeth.lmer.full	3	163.7375	172.0240
SelMeth.lmer.allfixed	8	160.7925	182.8899
SelMeth.lmer.fixed.3	6	160.3294	176.9024
SelMeth.lmer.fixed.2	5	168.8274	182.6383

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 217.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1448 -0.6998  0.5175  0.7452  2.6532
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08219  0.2867
##  Residual             0.29136  0.5398
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.48427    0.05962   41.67
```

```
InterpRes.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                               data=InterpRes.ratings.full1)
summary(InterpRes.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 203.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5878 -0.7454  0.1800  0.5704  2.6492
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06532  0.2556
##  Residual             0.25357  0.5036
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  3.56231    0.57808   6.162
## as.factor(Rater)2  3.44479    0.57805   5.959
## as.factor(Rater)3  3.00000    0.56471   5.313
## SemesterS19       -0.09219    0.12586  -0.733
```

```
## SexF              -0.80598    0.57679  -1.397
## SexM              -0.86368    0.57532  -1.501
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.977
## as.fctr(R)3  0.977  0.977
## SemesterS19  0.000  0.002  0.000
## SexF        -0.981 -0.980 -0.979 -0.092
## SexM        -0.981 -0.982 -0.982 -0.034  0.980
```

```
anova(InterpRes.lmer.full,InterpRes.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InterpRes.ratings.full1
## Models:
## InterpRes.lmer.full: Rating ~ 1 + (1 | Artifact)
## InterpRes.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                         npar    AIC    BIC   logLik deviance  Chisq Df
## InterpRes.lmer.full        3 220.09 228.38 -107.048   214.09
## InterpRes.lmer.allfixed    8 206.65 228.75  -95.326   190.65 23.444  5
##                         Pr(>Chisq)
## InterpRes.lmer.full
## InterpRes.lmer.allfixed  0.0002776 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
InterpRes.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                              data=InterpRes.ratings.full1)
summary(InterpRes.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 204.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5279 -0.7642  0.3371  0.6091  2.6242
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06792  0.2606
##  Residual             0.25235  0.5023
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1  2.73130    0.09741   28.040
## as.factor(Rater)2  2.61164    0.09634   27.108
```

```
## as.factor(Rater)3  2.18629    0.09634  22.693
## SemesterS19       -0.08543    0.12067  -0.708
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.197
## as.fctr(R)3  0.197  0.188
## SemesterS19 -0.388 -0.362 -0.362
```

```
anova(InterpRes.lmer.allfixed,InterpRes.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)

## Data: InterpRes.ratings.full1
## Models:
## InterpRes.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
## InterpRes.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                         npar    AIC    BIC  logLik deviance  Chisq Df
## InterpRes.lmer.fixed.3     6 205.15 221.72 -96.573   193.15
## InterpRes.lmer.allfixed    8 206.65 228.75 -95.326   190.65 2.4942  2
##                         Pr(>Chisq)
## InterpRes.lmer.fixed.3
## InterpRes.lmer.allfixed     0.2873
```

```
InterpRes.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                              data=InterpRes.ratings.full1)
summary(InterpRes.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 202.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5101 -0.7484  0.3763  0.6532  2.6479
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06471  0.2544
##  Residual             0.25381  0.5038
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.70517    0.08961   30.19
## as.factor(Rater)2  2.58701    0.08961   28.87
## as.factor(Rater)3  2.16116    0.08961   24.12
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.063
## as.fctr(R)3 0.063  0.063
```

	df	AIC	BIC
InterpRes.lmer.full	3	223.9031	232.1896
InterpRes.lmer.allfixed	8	219.9267	242.0241
InterpRes.lmer.fixed.3	6	216.5678	233.1409
InterpRes.lmer.fixed.2	5	212.6734	226.4843

```
anova(InterpRes.lmer.fixed.3,InterpRes.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: InterpRes.ratings.full1
## Models:
## InterpRes.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
## InterpRes.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InterpRes.lmer.fixed.2    5 203.66 217.47 -96.831   193.66
## InterpRes.lmer.fixed.3    6 205.15 221.72 -96.573   193.15 0.5168  1     0.4722
```

```
ICC.InterpRes.fixed = (0.06224)/(0.06224+0.25250)
ICC.InterpRes.fixed
```

```
## [1] 0.1977505
```

```
data.frame(AIC=AIC(InterpRes.lmer.full,InterpRes.lmer.allfixed,InterpRes.lmer.fixed.3,InterpRes.lmer.fi
           BIC(InterpRes.lmer.full,InterpRes.lmer.allfixed,InterpRes.lmer.fixed.3,InterpRes.lmer.fixed.
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```
VisOrg.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="VisOrg",]
VisOrg.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=VisOrg.ratings.full1)
summary(VisOrg.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: VisOrg.ratings.full1
##
## REML criterion at convergence: 226.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5918 -0.3789 -0.1632  0.4726  1.6322
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3092   0.5561
##  Residual             0.1588   0.3985
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44497    0.07063   34.62
```

```r
VisOrg.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                             data=VisOrg.ratings.full1)
summary(VisOrg.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: VisOrg.ratings.full1
##
## REML criterion at convergence: 220.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5498 -0.4041 -0.1828  0.4038  1.8478
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2797   0.5289
##  Residual             0.1494   0.3865
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   3.0958     0.6655   4.652
## as.factor(Rater)2   3.3658     0.6654   5.058
## as.factor(Rater)3   3.0000     0.6551   4.580
## SemesterS19        -0.2458     0.1546  -1.590
## SexF               -0.5513     0.6686  -0.825
## SexM               -0.7505     0.6665  -1.126
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.985
## as.fctr(R)3  0.984  0.984
## SemesterS19 -0.002  0.002  0.000
## SexF        -0.980 -0.980 -0.980 -0.102
## SexM        -0.982 -0.983 -0.983 -0.035  0.977
```

```r
anova(VisOrg.lmer.full,VisOrg.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings.full1
## Models:
## VisOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## VisOrg.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg.lmer.full        3 228.95 237.21 -111.47   222.95
## VisOrg.lmer.allfixed    8 224.30 246.33 -104.15   208.30 14.647  5    0.01198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
VisOrg.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                          data=VisOrg.ratings.full1)
summary(VisOrg.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: VisOrg.ratings.full1
##
## REML criterion at convergence: 222.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5250 -0.3885 -0.1463  0.4205  1.8365
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2894   0.5379
##  Residual             0.1462   0.3823
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.4426     0.1074  22.744
## as.factor(Rater)2   2.7096     0.1052  25.747
## as.factor(Rater)3   2.3553     0.1052  22.380
## SemesterS19        -0.1886     0.1477  -1.277
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.400
## as.fctr(R)3  0.400  0.392
## SemesterS19 -0.442 -0.423 -0.423
```

```r
anova(VisOrg.lmer.full,VisOrg.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings.full1
## Models:
## VisOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## VisOrg.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg.lmer.full       3 228.95 237.21 -111.47   222.95
## VisOrg.lmer.fixed.3    6 223.31 239.83 -105.66   211.31 11.639  3   0.008728 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
VisOrg.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                          data=VisOrg.ratings.full1)
summary(VisOrg.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: VisOrg.ratings.full1
##
## REML criterion at convergence: 221.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5008 -0.3334 -0.2599  0.4108  1.8726
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2937   0.5420
##  Residual             0.1454   0.3813
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                  Estimate Std. Error t value
## as.factor(Rater)1  2.38148    0.09652   24.67
## as.factor(Rater)2  2.65269    0.09558   27.75
## as.factor(Rater)3  2.29935    0.09558   24.06
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.265
## as.fctr(R)3 0.265  0.264
```

```
anova(VisOrg.lmer.full,VisOrg.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings.full1
## Models:
## VisOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## VisOrg.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg.lmer.full       3 228.95 237.21 -111.47   222.95
## VisOrg.lmer.fixed.2    5 222.97 236.74 -106.48   212.97 9.9784  2   0.006811 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ICC.VisOrg.fixed = (0.2907)/(0.2907+0.1467)
ICC.VisOrg.fixed
```

```
## [1] 0.6646091
```

```
data.frame(AIC=AIC(VisOrg.lmer.full,VisOrg.lmer.allfixed,VisOrg.lmer.fixed.3,VisOrg.lmer.fixed.2),
           BIC(VisOrg.lmer.full,VisOrg.lmer.allfixed,VisOrg.lmer.fixed.3,VisOrg.lmer.fixed.2))[,-3] %>%
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

	df	AIC	BIC
VisOrg.lmer.full	3	232.4172	240.6780
VisOrg.lmer.allfixed	8	236.2432	258.2719
VisOrg.lmer.fixed.3	6	234.1226	250.6441
VisOrg.lmer.fixed.2	5	231.7626	245.5305

```
TxtOrg.ratings.full1 <- tall.full.nonmissing[tall.full.nonmissing$Rubric=="TxtOrg",]
TxtOrg.lmer.full <- lmer(Rating ~ 1 + (1|Artifact), data=TxtOrg.ratings.full1)
summary(TxtOrg.lmer.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings.full1
##
## REML criterion at convergence: 249
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3638 -0.7641  0.3836  0.5278  2.4094
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09145  0.3024
##  Residual             0.39503  0.6285
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.59144    0.06764   38.31
```

```
TxtOrg.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),
                            data=TxtOrg.ratings.full1)
summary(TxtOrg.lmer.allfixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##    Data: TxtOrg.ratings.full1
##
## REML criterion at convergence: 249
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3911 -0.6381  0.2279  0.5618  2.3900
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06472  0.2544
##  Residual             0.41037  0.6406
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
```

```
##                     Estimate Std. Error t value
## as.factor(Rater)1   3.3368     0.7061   4.726
## as.factor(Rater)2   3.1577     0.7061   4.472
## as.factor(Rater)3   3.0000     0.6893   4.352
## SemesterS19        -0.2090     0.1516  -1.379
## SexF               -0.4755     0.7039  -0.676
## SexM               -0.5634     0.7022  -0.802
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3 SmsS19 SexF
## as.fctr(R)2  0.977
## as.fctr(R)3  0.976  0.976
## SemesterS19  0.000  0.002  0.000
## SexF        -0.981 -0.980 -0.979 -0.090
## SexM        -0.981 -0.982 -0.982 -0.034  0.980
```

```
anova(TxtOrg.lmer.full,TxtOrg.lmer.allfixed)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: TxtOrg.ratings.full1
## Models:
## TxtOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## TxtOrg.lmer.allfixed: Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg.lmer.full        3 251.45 259.74 -122.73   245.45
## TxtOrg.lmer.allfixed    8 254.10 276.20 -119.05   238.10 7.3492  5     0.1959
```

```
TxtOrg.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact),
                           data=TxtOrg.ratings.full1)
summary(TxtOrg.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: TxtOrg.ratings.full1
##
## REML criterion at convergence: 249
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4379 -0.6578  0.2435  0.5747  2.2764
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.0720   0.2683
##  Residual             0.3994   0.6320
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1   2.8184     0.1184  23.809
## as.factor(Rater)2   2.6358     0.1171  22.513
```

```
## as.factor(Rater)3   2.4907     0.1171  21.273
## SemesterS19        -0.1886     0.1449  -1.302
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.177
## as.fctr(R)3  0.177  0.168
## SemesterS19 -0.382 -0.356 -0.356
```

```
anova(TxtOrg.lmer.full,TxtOrg.lmer.fixed.3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: TxtOrg.ratings.full1
## Models:
## TxtOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## TxtOrg.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg.lmer.full       3 251.45 259.74 -122.73   245.45
## TxtOrg.lmer.fixed.3    6 251.12 267.70 -119.56   239.12 6.3299  3    0.09662 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TxtOrg.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                           data=TxtOrg.ratings.full1)
summary(TxtOrg.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: TxtOrg.ratings.full1
##
## REML criterion at convergence: 248.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3539 -0.5816  0.3207  0.5579  2.1098
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07478  0.2735
##  Residual             0.39969  0.6322
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.7593     0.1097   25.14
## as.factor(Rater)2   2.5813     0.1097   23.52
## as.factor(Rater)3   2.4365     0.1097   22.20
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.049
## as.fctr(R)3 0.049  0.049
```

	df	AIC	BIC
TxtOrg.lmer.full	3	255.0106	263.2972
TxtOrg.lmer.allfixed	8	265.0165	287.1138
TxtOrg.lmer.fixed.3	6	260.9890	277.5620
TxtOrg.lmer.fixed.2	5	258.6555	272.4663

```
anova(TxtOrg.lmer.full,TxtOrg.lmer.fixed.2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: TxtOrg.ratings.full1
## Models:
## TxtOrg.lmer.full: Rating ~ 1 + (1 | Artifact)
## TxtOrg.lmer.fixed.2: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg.lmer.full       3 251.45 259.74 -122.73   245.45
## TxtOrg.lmer.fixed.2    5 250.86 264.68 -120.43   240.86 4.5892  2     0.1008
```

```
ICC.TxtOrg.fixed = (0.09371)/(0.09371+0.39573)
ICC.TxtOrg.fixed
```

```
## [1] 0.1914637
```

```
data.frame(AIC=AIC(TxtOrg.lmer.full,TxtOrg.lmer.allfixed,TxtOrg.lmer.fixed.3,TxtOrg.lmer.fixed.2),
           BIC(TxtOrg.lmer.full,TxtOrg.lmer.allfixed,TxtOrg.lmer.fixed.3,TxtOrg.lmer.fixed.2))[,-3] %>%
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

Random Effects per Rubric (Full Data)

After finding the best fixed effect model for each rubric we tested some different random effects. We used AIC/BIC to decide on the best models.

```
SelMeth.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester +
                              (1|Artifact), data=SelMeth.ratings.full1)
summary(SelMeth.lmer.fixed.3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 148.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.07037 -0.37872 -0.09732  0.22451  2.50908
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09505  0.3083
##  Residual             0.11002  0.3317
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.25854    0.07612  29.671
## as.factor(Rater)2  2.23392    0.07532  29.658
## as.factor(Rater)3  2.06017    0.07532  27.351
## SemesterS19       -0.37273    0.09954  -3.744
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.288
## as.fctr(R)3  0.288  0.280
## SemesterS19 -0.412 -0.390 -0.390
```

```
SelMeth.lmer.int <-lmer(Rating ~ -1 + as.factor(Rater)*Semester - Semester +
                          (1|Artifact), data=SelMeth.ratings.full1)
summary(SelMeth.lmer.int)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) * Semester - Semester + (1 | Artifact)
##    Data: SelMeth.ratings.full1
##
## REML criterion at convergence: 148.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2323 -0.3602 -0.1860  0.1609  2.3356
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09077  0.3013
##  Residual             0.11181  0.3344
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                               Estimate Std. Error t value
## as.factor(Rater)1              2.21909    0.08286  26.782
## as.factor(Rater)2              2.21824    0.08152  27.210
## as.factor(Rater)3              2.11270    0.08152  25.915
## as.factor(Rater)1:SemesterS19 -0.24056    0.15045  -1.599
## as.factor(Rater)2:SemesterS19 -0.31573    0.15403  -2.050
## as.factor(Rater)3:SemesterS19 -0.56834    0.15403  -3.690
##
## Correlation of Fixed Effects:
##             as.(R)1 as.(R)2 as.(R)3 a.(R)1: a.(R)2:
## as.fctr(R)2  0.158
## as.fctr(R)3  0.158   0.156
## a.(R)1:SS19 -0.551  -0.087  -0.087
## a.(R)2:SS19 -0.084  -0.529  -0.083   0.126
## a.(R)3:SS19 -0.084  -0.083  -0.529   0.126   0.128
```

```r
anova(SelMeth.lmer.fixed.3,SelMeth.lmer.int)
```

```
## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings.full1
## Models:
## SelMeth.lmer.fixed.3: Rating ~ -1 + as.factor(Rater) + Semester + (1 | Artifact)
## SelMeth.lmer.int: Rating ~ -1 + as.factor(Rater) * Semester - Semester + (1 | Artifact)
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth.lmer.fixed.3    6 146.92 163.49 -67.461   134.92
## SelMeth.lmer.int        8 147.93 170.02 -65.963   131.93 2.9948  2     0.2237
```

```r
# SelMeth.lmer.ran1 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact) +
#(Semester|Artifact), data=SelMeth.ratings.full1)
# summary(SelMeth.lmer.ran1)
# SelMeth.lmer.ran2 <- lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact) +
#(as.factor(Rater)|Artifact), data=SelMeth.ratings.full1)
# summary(SelMeth.lmer.ran2)


InterpRes.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                             data=InterpRes.ratings.full1)
summary(InterpRes.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 202.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5101 -0.7484  0.3763  0.6532  2.6479
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06471  0.2544
##  Residual             0.25381  0.5038
## Number of obs: 117, groups:  Artifact, 91
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.70517    0.08961   30.19
## as.factor(Rater)2  2.58701    0.08961   28.87
## as.factor(Rater)3  2.16116    0.08961   24.12
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.063
## as.fctr(R)3 0.063  0.063
```

```r
# InterpRes.lmer.ran <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact)+
#(as.factor(Rater)|Artifact), data=InterpRes.ratings.full1)
# summary(InterpRes.lmer.ran)



VisOrg.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact),
                           data=VisOrg.ratings.full1)
summary(VisOrg.lmer.fixed.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + (1 | Artifact)
##    Data: VisOrg.ratings.full1
##
## REML criterion at convergence: 221.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5008 -0.3334 -0.2599  0.4108  1.8726
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2937   0.5420
##  Residual             0.1454   0.3813
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.38148    0.09652   24.67
## as.factor(Rater)2  2.65269    0.09558   27.75
## as.factor(Rater)3  2.29935    0.09558   24.06
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.265
## as.fctr(R)3 0.265  0.264
```

```r
# VisOrg.lmer.ran <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact) +
#(as.factor(Rater)|Artifact), data=VisOrg.ratings.full1)
# summary(VisOrg.lmer.ran)
```

This method is flawed because it doesn't let you directly examine interactions with Rubric. To combat this we used the full data set without subsetting by each rubric individually.

Fixed Effects Overall (Full Data)

We manually selected fixed effects, random effects, and considered interactions to find the best combined model. We used likelihood ratio tests and AIC/BIC to compare models.

```r
null.lmer <- lmer(Rating ~ 1 + (0 + Rubric | Artifact),
                  data=tall.full.nonmissing, REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00449929 (tol = 0.002, component 1)
```

```r
summary(null.lmer)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ 1 + (0 + Rubric | Artifact)
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1537.2   1678.3   -738.6   1477.2      787
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0239 -0.4975 -0.0754  0.5217  3.7844
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.6470   0.8044
##           RubricInitEDA   0.3762   0.6133   0.27
##           RubricInterpRes 0.2510   0.5010   0.01 0.79
##           RubricRsrchQ    0.1717   0.4143   0.40 0.50 0.74
##           RubricSelMeth   0.1020   0.3194   0.58 0.38 0.41 0.28
##           RubricTxtOrg    0.3930   0.6269   0.03 0.69 0.80 0.64 0.24
##           RubricVisOrg    0.3136   0.5600   0.19 0.78 0.77 0.60 0.30 0.79
##  Residual                 0.1942   0.4406
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.24700    0.04026   55.81
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00449929 (tol = 0.002, component 1)
```

```r
all.fixed.lmer <- lmer(Rating ~ -1 + as.factor(Rater) + Repeated + Semester +
                         Sex + Rubric +
                         (0 + Rubric | Artifact),
                       data=tall.full.nonmissing, REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00241128 (tol = 0.002, component 1)
```

```r
summary(all.fixed.lmer)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Repeated + Semester + Sex +
##     Rubric + (0 + Rubric | Artifact)
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1476.0   1673.6   -696.0   1392.0      775
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1325 -0.5090 -0.0225  0.5268  3.8193
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.53612  0.7322
##           RubricInitEDA   0.33886  0.5821   0.46
##           RubricInterpRes 0.16551  0.4068   0.22 0.75
##           RubricRsrchQ    0.15995  0.3999   0.58 0.43 0.70
##           RubricSelMeth   0.06133  0.2476   0.38 0.59 0.73 0.38
##           RubricTxtOrg    0.25214  0.5021   0.33 0.61 0.70 0.55 0.66
##           RubricVisOrg    0.24604  0.4960   0.34 0.73 0.67 0.51 0.39 0.76
##  Residual                 0.18654  0.4319
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.82145    0.37757   7.473
## as.factor(Rater)2  2.82396    0.37757   7.479
## as.factor(Rater)3  2.64505    0.37364   7.079
## Repeated          -0.07415    0.09499  -0.781
## SemesterS19       -0.17493    0.08512  -2.055
## SexF              -0.80312    0.37248  -2.156
## SexM              -0.79350    0.37151  -2.136
## RubricInitEDA      0.54129    0.09435   5.737
## RubricInterpRes    0.58085    0.09945   5.841
## RubricRsrchQ       0.45602    0.08631   5.283
## RubricSelMeth      0.16293    0.09290   1.754
## RubricTxtOrg       0.68580    0.09816   6.987
## RubricVisOrg       0.52419    0.09771   5.365

##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00241128 (tol = 0.002, component 1)
```

```
fixed.lmer <-lmer(Rating ~ -1 + as.factor(Rater) + Rubric + Semester +
                    (0 + Rubric | Artifact),
                  data=tall.full.nonmissing, REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00397028 (tol = 0.002, component 1)
```

```
summary(fixed.lmer)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) + Rubric + Semester + (0 + Rubric |
##     Artifact)
```

```
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1475.2   1658.7   -698.6   1397.2      778
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1325 -0.5135 -0.0138  0.5241  3.7919
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.55283  0.7435
##           RubricInitEDA   0.34354  0.5861   0.47
##           RubricInterpRes 0.16745  0.4092   0.25 0.75
##           RubricRsrchQ    0.16679  0.4084   0.60 0.45 0.71
##           RubricSelMeth   0.06681  0.2585   0.43 0.62 0.75 0.43
##           RubricTxtOrg    0.24872  0.4987   0.34 0.62 0.70 0.56 0.66
##           RubricVisOrg    0.25421  0.5042   0.36 0.74 0.69 0.53 0.43 0.76
##  Residual                 0.18755  0.4331
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.02251    0.09811  20.614
## as.factor(Rater)2  2.02129    0.09791  20.645
## as.factor(Rater)3  1.85313    0.09788  18.933
## RubricInitEDA      0.54178    0.09436   5.742
## RubricInterpRes    0.58019    0.09944   5.835
## RubricRsrchQ       0.45324    0.08622   5.257
## RubricSelMeth      0.15648    0.09223   1.697
## RubricTxtOrg       0.68602    0.09819   6.987
## RubricVisOrg       0.52274    0.09765   5.353
## SemesterS19       -0.18421    0.08299  -2.220
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO
## as.fctr(R)2  0.845
## as.fctr(R)3  0.845  0.844
## RubrcIntEDA -0.611 -0.612 -0.612
## RbrcIntrpRs -0.733 -0.735 -0.735  0.734
## RubrcRsrchQ -0.700 -0.702 -0.701  0.588  0.756
## RubricSlMth -0.778 -0.780 -0.780  0.663  0.778  0.689
## RubrcTxtOrg -0.680 -0.682 -0.682  0.676  0.752  0.684  0.728
## RubricVsOrg -0.674 -0.676 -0.676  0.716  0.745  0.668  0.682  0.752
## SemesterS19 -0.265 -0.257 -0.257 -0.001  0.000  0.002  0.006 -0.001  0.000
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00397028 (tol = 0.002, component 1)
```

```r
int.lmer <- lmer(Rating ~ -1 + as.factor(Rater)*Rubric + as.factor(Rater)*Semester +
                   Semester*Rubric + (0 + Rubric | Artifact),
                 data=tall.full.nonmissing, REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
summary(int.lmer)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) * Rubric + as.factor(Rater) *
##     Semester + Semester * Rubric + (0 + Rubric | Artifact)
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1468.9   1746.5   -675.4   1350.9      758
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9554 -0.5299 -0.0551  0.5110  3.6730
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.49618  0.7044
##           RubricInitEDA   0.34129  0.5842   0.45
##           RubricInterpRes 0.14801  0.3847   0.36 0.81
##           RubricRsrchQ    0.16293  0.4037   0.65 0.43 0.72
##           RubricSelMeth   0.06746  0.2597   0.46 0.64 0.80 0.49
##           RubricTxtOrg    0.25193  0.5019   0.42 0.64 0.67 0.57 0.63
##           RubricVisOrg    0.24821  0.4982   0.35 0.72 0.68 0.55 0.39 0.77
##  Residual                 0.17741  0.4212
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                                    Estimate Std. Error t value
## as.factor(Rater)1                  1.726094   0.128281  13.456
## as.factor(Rater)2                  2.100697   0.127341  16.497
## as.factor(Rater)3                  1.980323   0.127188  15.570
## RubricInitEDA                      0.727530   0.148403   4.902
## RubricInterpRes                    1.008786   0.146269   6.897
## RubricRsrchQ                       0.683477   0.133029   5.138
## RubricSelMeth                      0.506404   0.140576   3.602
## RubricTxtOrg                       1.076900   0.147138   7.319
## RubricVisOrg                       0.710716   0.151573   4.689
## SemesterS19                       -0.032069   0.197955  -0.162
## as.factor(Rater)2:RubricInitEDA   -0.303363   0.169180  -1.793
## as.factor(Rater)3:RubricInitEDA   -0.302114   0.168697  -1.791
## as.factor(Rater)2:RubricInterpRes -0.532087   0.166853  -3.189
## as.factor(Rater)3:RubricInterpRes -0.748540   0.166446  -4.497
## as.factor(Rater)2:RubricRsrchQ    -0.490014   0.157534  -3.111
## as.factor(Rater)3:RubricRsrchQ    -0.364829   0.156944  -2.325
## as.factor(Rater)2:RubricSelMeth   -0.401483   0.161059  -2.493
## as.factor(Rater)3:RubricSelMeth   -0.403398   0.160600  -2.512
## as.factor(Rater)2:RubricTxtOrg    -0.582740   0.168309  -3.462
## as.factor(Rater)3:RubricTxtOrg    -0.495779   0.167855  -2.954
## as.factor(Rater)2:RubricVisOrg    -0.144614   0.171085  -0.845
## as.factor(Rater)3:RubricVisOrg    -0.335860   0.170673  -1.968
## as.factor(Rater)2:SemesterS19     -0.045671   0.122190  -0.374
## as.factor(Rater)3:SemesterS19     -0.127593   0.122184  -1.044
## RubricInitEDA:SemesterS19          0.052587   0.202557   0.260
```

```
## RubricInterpRes:SemesterS19        0.002651   0.199713   0.013
## RubricRsrchQ:SemesterS19           0.190756   0.176918   1.078
## RubricSelMeth:SemesterS19         -0.273938   0.191690  -1.429
## RubricTxtOrg:SemesterS19          -0.099038   0.200351  -0.494
## RubricVisOrg:SemesterS19          -0.099649   0.206890  -0.482
```

```
##
## Correlation matrix not shown by default, as p = 30 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
int.lmer2 <- lmer(Rating ~ -1 + as.factor(Rater)*Rubric  + Semester +
                    (0 + Rubric | Artifact),
                  data=tall.full.nonmissing, REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(int.lmer2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) * Rubric + Semester + (0 + Rubric |
##     Artifact)
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1465.8   1705.8   -681.9   1363.8      766
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0194 -0.5240 -0.0435  0.4901  3.6094
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.49956  0.7068
##           RubricInitEDA   0.34758  0.5896   0.45
##           RubricInterpRes 0.15219  0.3901   0.38 0.82
##           RubricRsrchQ    0.17880  0.4228   0.64 0.45 0.73
##           RubricSelMeth   0.07013  0.2648   0.45 0.61 0.76 0.40
##           RubricTxtOrg    0.25289  0.5029   0.43 0.64 0.71 0.54 0.62
##           RubricVisOrg    0.25066  0.5007   0.35 0.72 0.69 0.52 0.40 0.78
##  Residual                 0.17963  0.4238
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                                    Estimate Std. Error t value
## as.factor(Rater)1                   1.77121    0.11659  15.192
## as.factor(Rater)2                   2.13565    0.11734  18.200
## as.factor(Rater)3                   1.99549    0.11648  17.131
## RubricInitEDA                       0.74530    0.13470   5.533
```

```
## RubricInterpRes                    1.01082    0.13257   7.625
## RubricRsrchQ                       0.74764    0.12232   6.112
## RubricSelMeth                      0.42395    0.12826   3.305
## RubricTxtOrg                       1.04545    0.13312   7.853
## RubricVisOrg                       0.68095    0.13728   4.960
## SemesterS19                       -0.19006    0.08256  -2.302
## as.factor(Rater)2:RubricInitEDA   -0.30798    0.16996  -1.812
## as.factor(Rater)3:RubricInitEDA   -0.30478    0.16951  -1.798
## as.factor(Rater)2:RubricInterpRes -0.53717    0.16737  -3.209
## as.factor(Rater)3:RubricInterpRes -0.75227    0.16700  -4.505
## as.factor(Rater)2:RubricRsrchQ    -0.50208    0.15907  -3.156
## as.factor(Rater)3:RubricRsrchQ    -0.37493    0.15853  -2.365
## as.factor(Rater)2:RubricSelMeth   -0.39767    0.16207  -2.454
## as.factor(Rater)3:RubricSelMeth   -0.40393    0.16165  -2.499
## as.factor(Rater)2:RubricTxtOrg    -0.58069    0.16850  -3.446
## as.factor(Rater)3:RubricTxtOrg    -0.49476    0.16807  -2.944
## as.factor(Rater)2:RubricVisOrg    -0.14481    0.17174  -0.843
## as.factor(Rater)3:RubricVisOrg    -0.33756    0.17136  -1.970
```

```
##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
anova(fit.lmer,fixed.lmer, int.lmer,int.lmer2)
```

```
## Error in anova(fit.lmer, fixed.lmer, int.lmer, int.lmer2): object 'fit.lmer' not found
```

```
int.ran.lmer <- lmer(Rating ~ -1 + as.factor(Rater)*Rubric + Semester +
                       (0 + Rubric | Artifact) + (0 + as.factor(Rater) | Artifact),
                     data=tall.full.nonmissing, REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(int.ran.lmer)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Rating ~ -1 + as.factor(Rater) * Rubric + Semester + (0 + Rubric |
##     Artifact) + (0 + as.factor(Rater) | Artifact)
##    Data: tall.full.nonmissing
##
##      AIC      BIC   logLik deviance df.resid
##   1425.9   1694.1   -655.9   1311.9      760
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.11798 -0.46601 -0.03293  0.45812  2.80235
##
```

```
## Random effects:
##  Groups     Name               Variance Std.Dev. Corr
##  Artifact   RubricCritDes      0.48646  0.6975
##             RubricInitEDA      0.30561  0.5528    0.32
##             RubricInterpRes    0.09719  0.3118    0.14  0.67
##             RubricRsrchQ       0.17427  0.4175    0.50  0.19  0.53
##             RubricSelMeth      0.03665  0.1915    0.15  0.21  0.36 -0.25
##             RubricTxtOrg       0.23837  0.4882    0.26  0.43  0.34  0.29  0.18
##             RubricVisOrg       0.22284  0.4721    0.17  0.50  0.43  0.27 -0.18
##  Artifact.1 as.factor(Rater)1 0.01378  0.1174
##             as.factor(Rater)2 0.11267  0.3357   -0.38
##             as.factor(Rater)3 0.10415  0.3227    0.41  0.69
##  Residual                     0.12935  0.3597
##
##
##
##
##
##
##
##   0.52
##
##
##
##
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                                    Estimate Std. Error t value
## as.factor(Rater)1                   1.76561    0.11236  15.714
## as.factor(Rater)2                   2.13223    0.12124  17.587
## as.factor(Rater)3                   1.97690    0.11767  16.800
## RubricInitEDA                       0.73684    0.12781   5.765
## RubricInterpRes                     0.98790    0.12552   7.871
## RubricRsrchQ                        0.72249    0.11598   6.230
## RubricSelMeth                       0.40687    0.12254   3.320
## RubricTxtOrg                        1.01150    0.12792   7.907
## RubricVisOrg                        0.65032    0.13127   4.954
## SemesterS19                        -0.16385    0.07571  -2.164
## as.factor(Rater)2:RubricInitEDA   -0.29940    0.15331  -1.953
## as.factor(Rater)3:RubricInitEDA   -0.30172    0.15297  -1.972
## as.factor(Rater)2:RubricInterpRes -0.51213    0.15066  -3.399
## as.factor(Rater)3:RubricInterpRes -0.71448    0.15022  -4.756
## as.factor(Rater)2:RubricRsrchQ    -0.48663    0.14459  -3.366
## as.factor(Rater)3:RubricRsrchQ    -0.32533    0.14400  -2.259
## as.factor(Rater)2:RubricSelMeth   -0.38565    0.14750  -2.615
## as.factor(Rater)3:RubricSelMeth   -0.37882    0.14628  -2.590
## as.factor(Rater)2:RubricTxtOrg    -0.54980    0.15372  -3.577
## as.factor(Rater)3:RubricTxtOrg    -0.45239    0.15338  -2.949
## as.factor(Rater)2:RubricVisOrg    -0.10368    0.15572  -0.666
## as.factor(Rater)3:RubricVisOrg    -0.27691    0.15538  -1.782

##
## Correlation matrix not shown by default, as p = 22 > 12.
```

	df	AIC	BIC
null.lmer	30	1537.159	1678.328
all.fixed.lmer	42	1475.953	1673.590
fixed.lmer	39	1475.166	1658.686
int.lmer	59	1468.871	1746.503
int.lmer2	51	1465.827	1705.814
int.ran.lmer	57	1425.889	1694.110

```
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

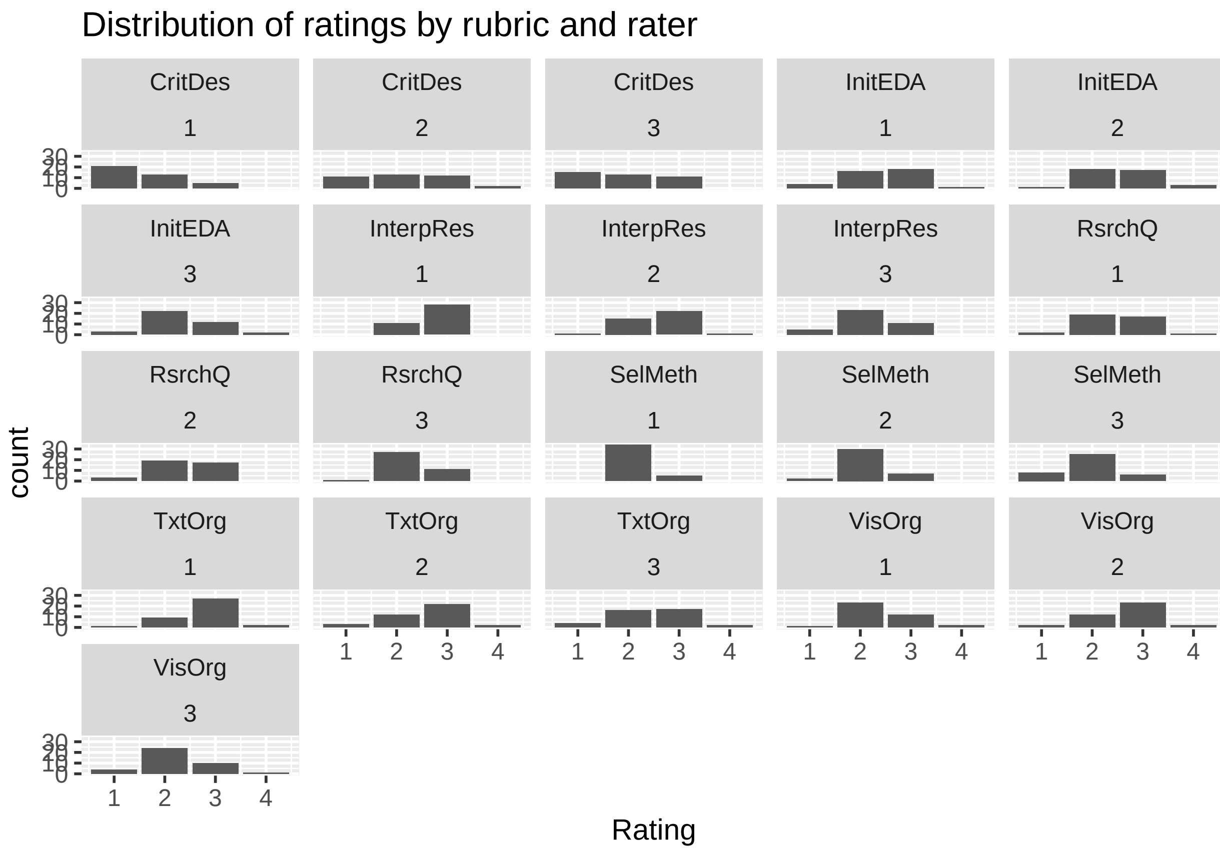
```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
data.frame(AIC=AIC(null.lmer,all.fixed.lmer,fixed.lmer,int.lmer, int.lmer2,int.ran.lmer),
           BIC(null.lmer,all.fixed.lmer,fixed.lmer,int.lmer,int.lmer2, int.ran.lmer))[,-3] %>%
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

Appendix 4. Extra Exploratory Data Analysis

To complete the analysis we performed some additional EDA using Semester and Sex to gather any other conclusions we felt weren't covered by the models.

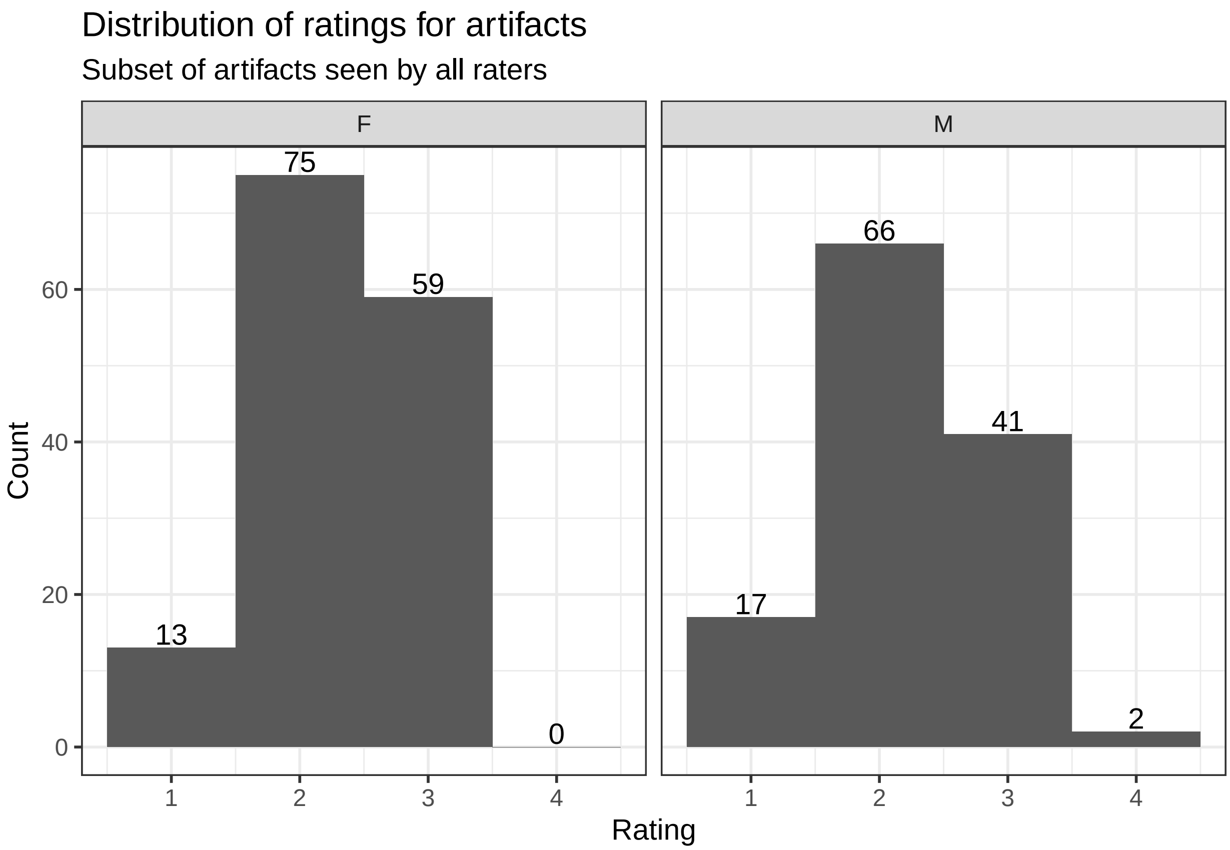
```
ggplot(tall.full.nonmissing, aes(x=Rating)) +
  geom_bar() +
  labs(title = "Distribution of ratings by rubric and rater")+
  facet_wrap( ~ Rubric + Rater)
```



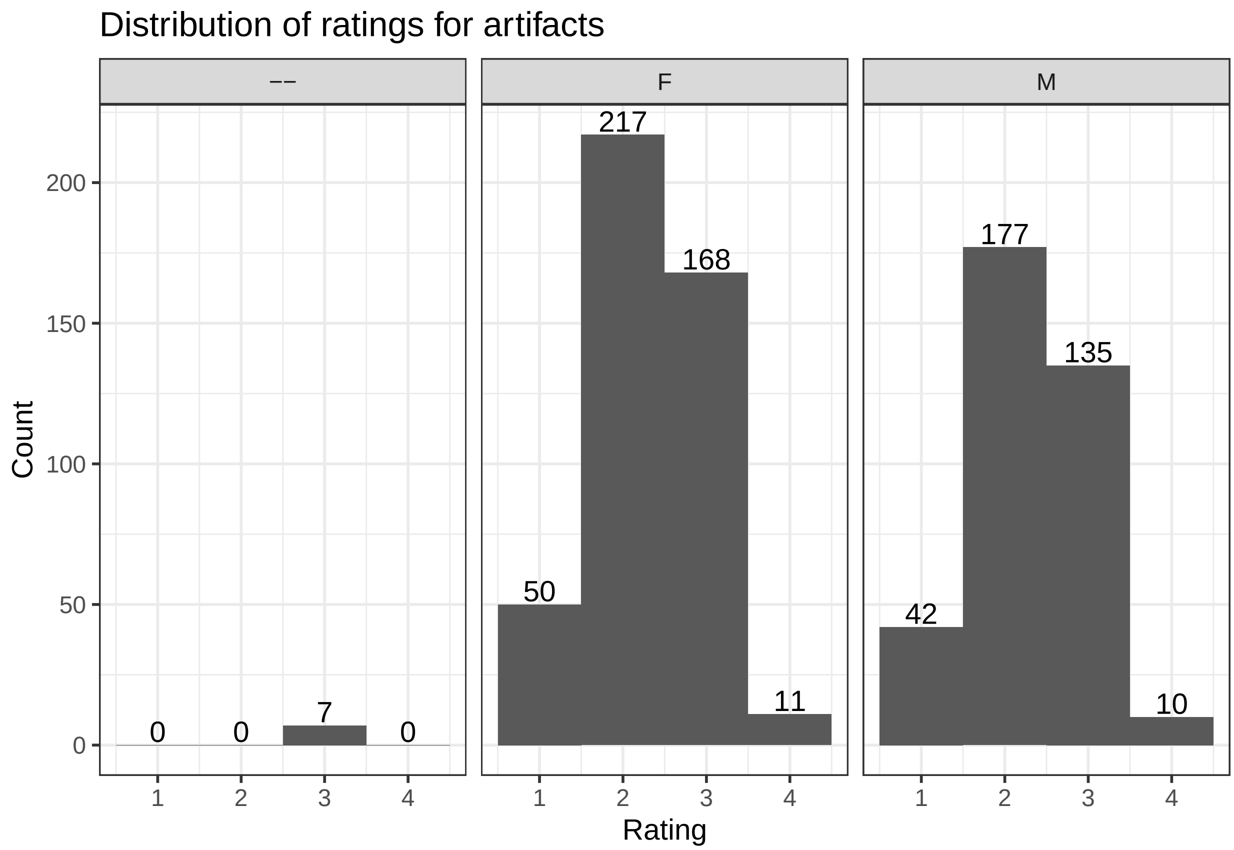

```r
ggsave("big.png")
```

```
## Saving 6.5 x 4.5 in image
```

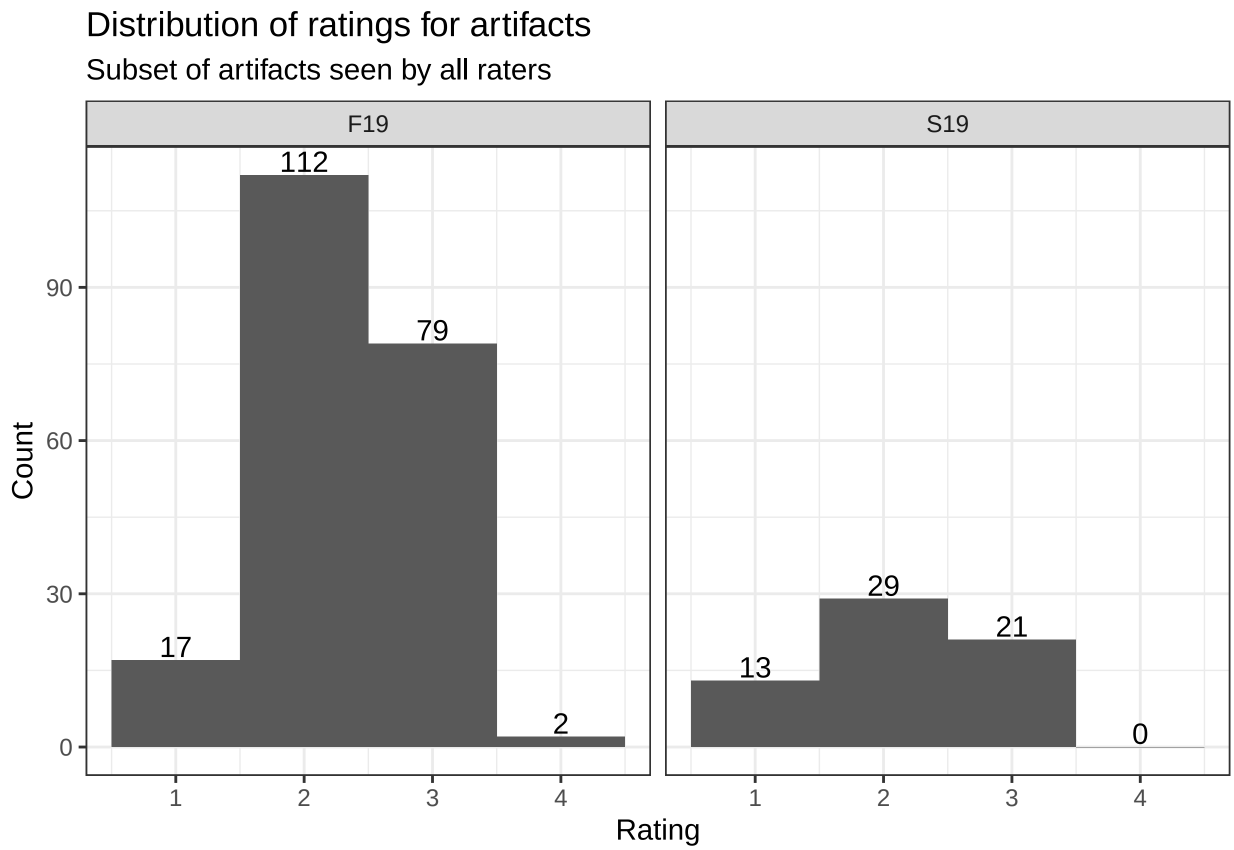
```r
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",
       subtitle = "Subset of artifacts seen by all raters") +
  theme_bw() +
  facet_wrap(~ Sex)
```




```
tall.full.nonmissing %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw() +
  facet_wrap(~ Sex)
```



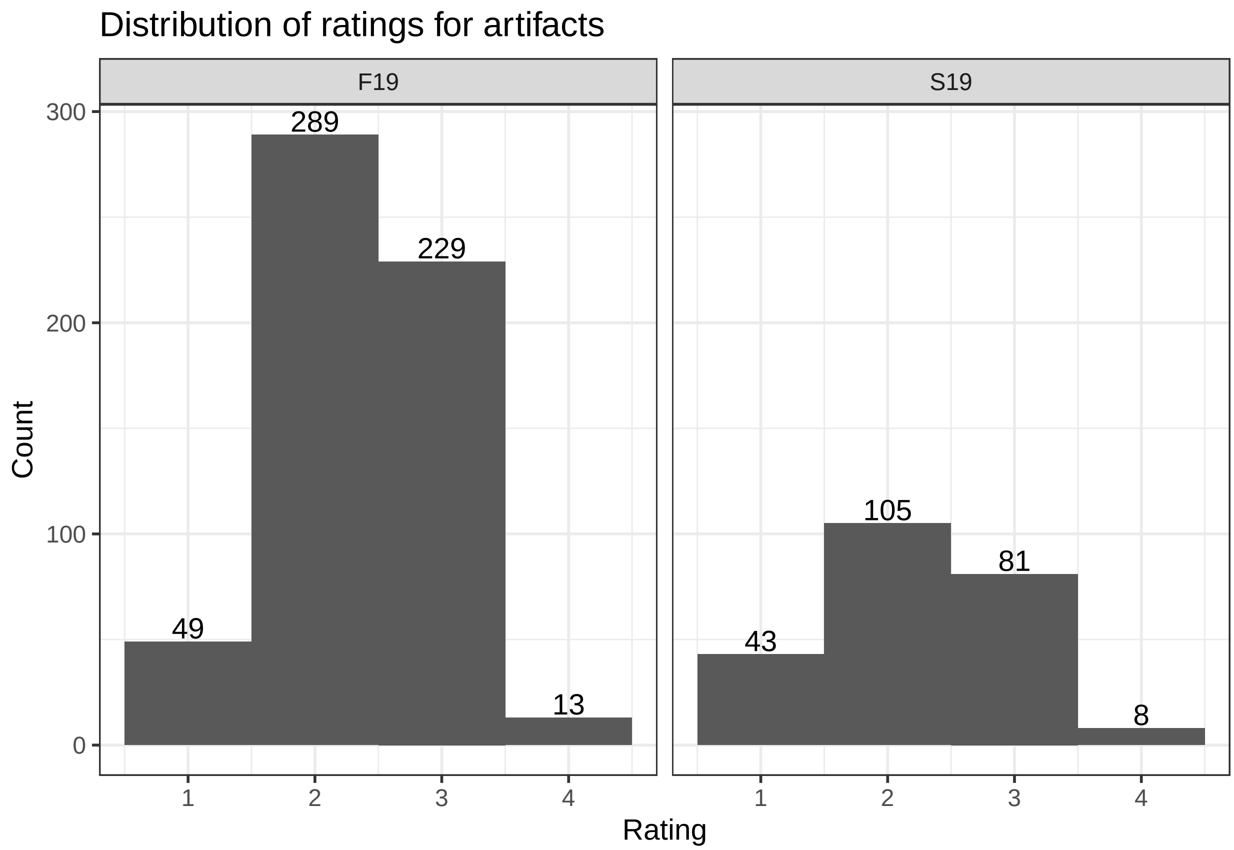

```r
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",
       subtitle = "Subset of artifacts seen by all raters") +
  theme_bw() +
  facet_wrap(~ Semester)
```




```
ggsave("Semester1.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
tall.full.nonmissing %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw() +
  facet_wrap(~ Semester)
```



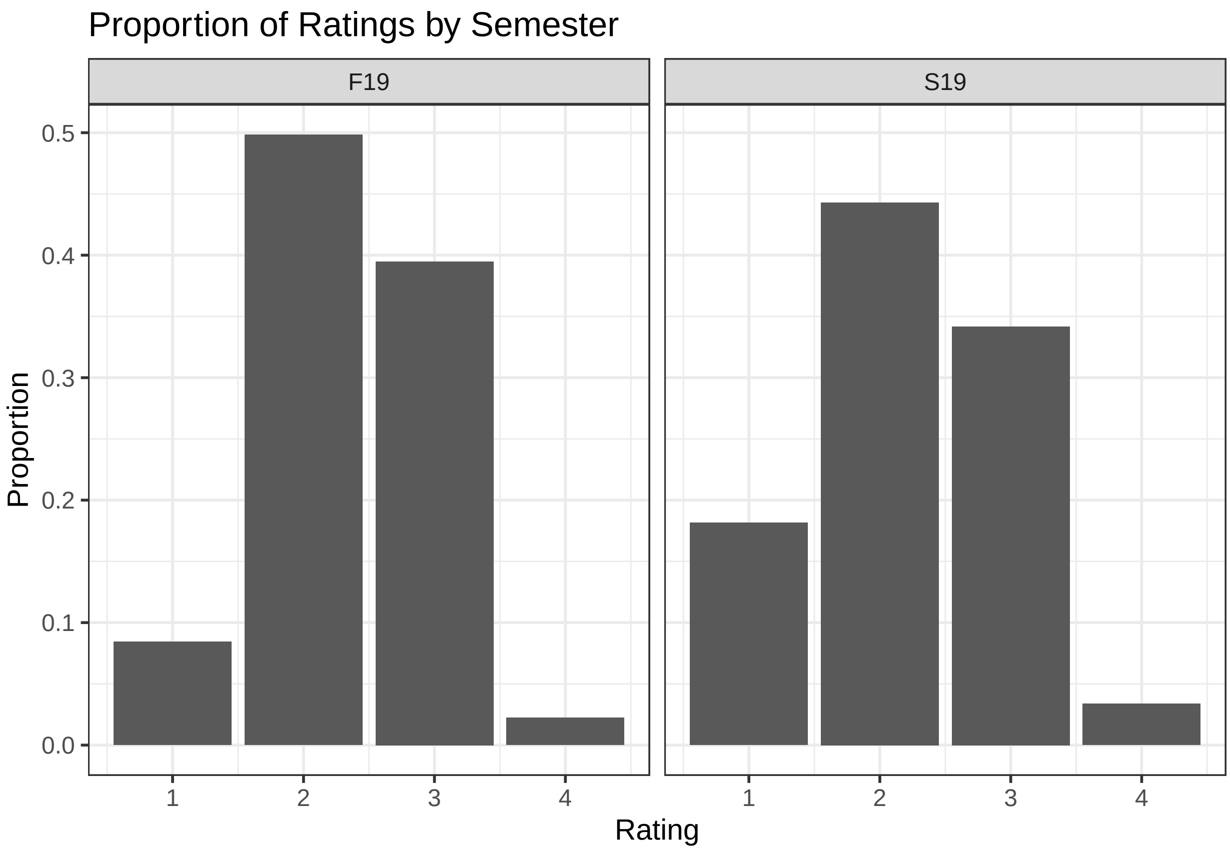

```
#Calculating proportions
prop <- tall.full.nonmissing %>%
  dplyr::select(Semester,Rating)
prop <- count(prop)

propf <- prop %>%
  filter(Semester == "F19") %>%
  mutate(prop = freq/sum(freq))

props <- prop %>%
  filter(Semester == "S19") %>%
  mutate(prop = freq/sum(freq))

propc <- rbind(propf,props)

ggplot(propc,aes(x = Rating, y= prop)) +
    geom_col() +
  labs(y = "Proportion", x = "Rating", title= "Proportion of Ratings by Semester") +
   theme_bw() +
  facet_wrap(~ Semester)
```




```
ggsave("prop.png")
```

```
## Saving 6.5 x 4.5 in image
```