(COOL TITLE I HAVEN'T THOUGHT OF YET)

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 requirements 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.

(WILL FINISH FOR FINAL DRAFT)

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 requirements to allow students to broaden their course of study. The Dietrich College at Carnegie Mellon University 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 labelled 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.

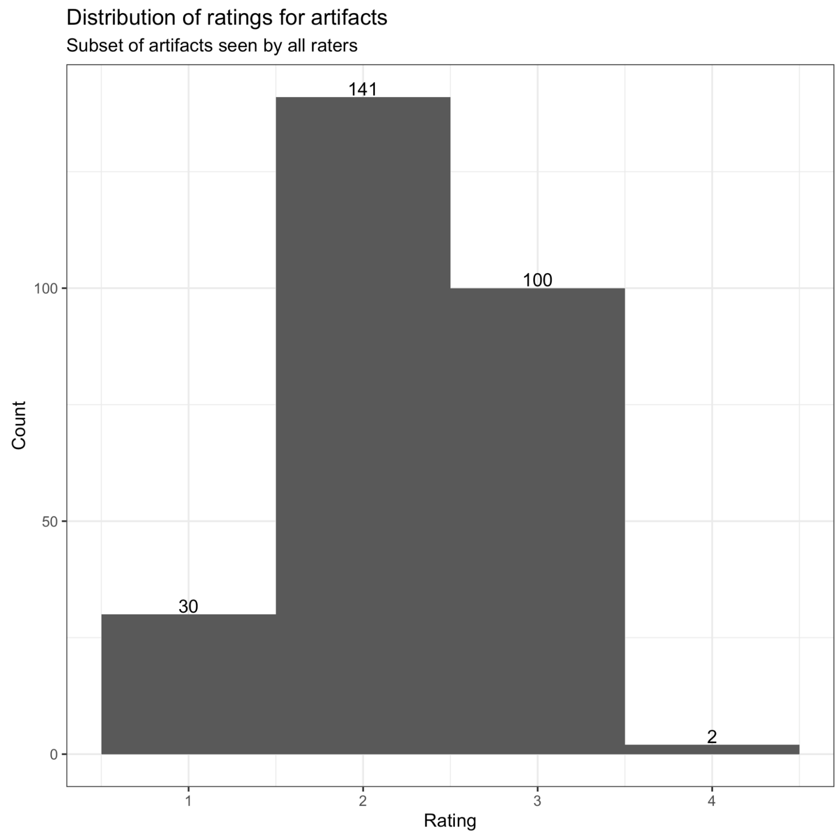


Figure 1: Distribution of ratings for artifacts seen by all three raters.

3 Methods

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

To begin answering the research questions, we first loaded 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. After this 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.

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. We are not able to do the same for the percent exact agreements because not all raters rated every artifact.

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.

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 models.

4 Results

(1) Distributions of Ratings

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. 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. 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. 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.

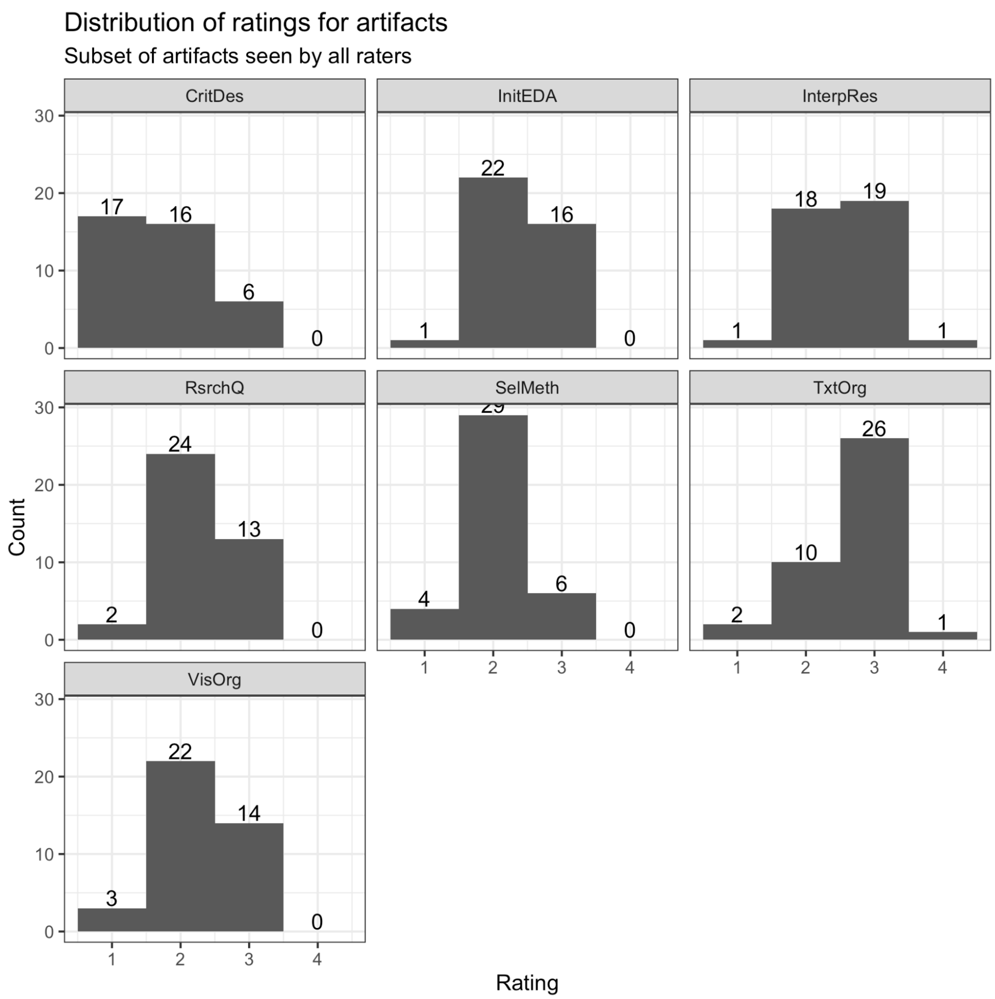


Figure 2: Distribution of ratings by rubric for subset of artifacts seen by all raters.

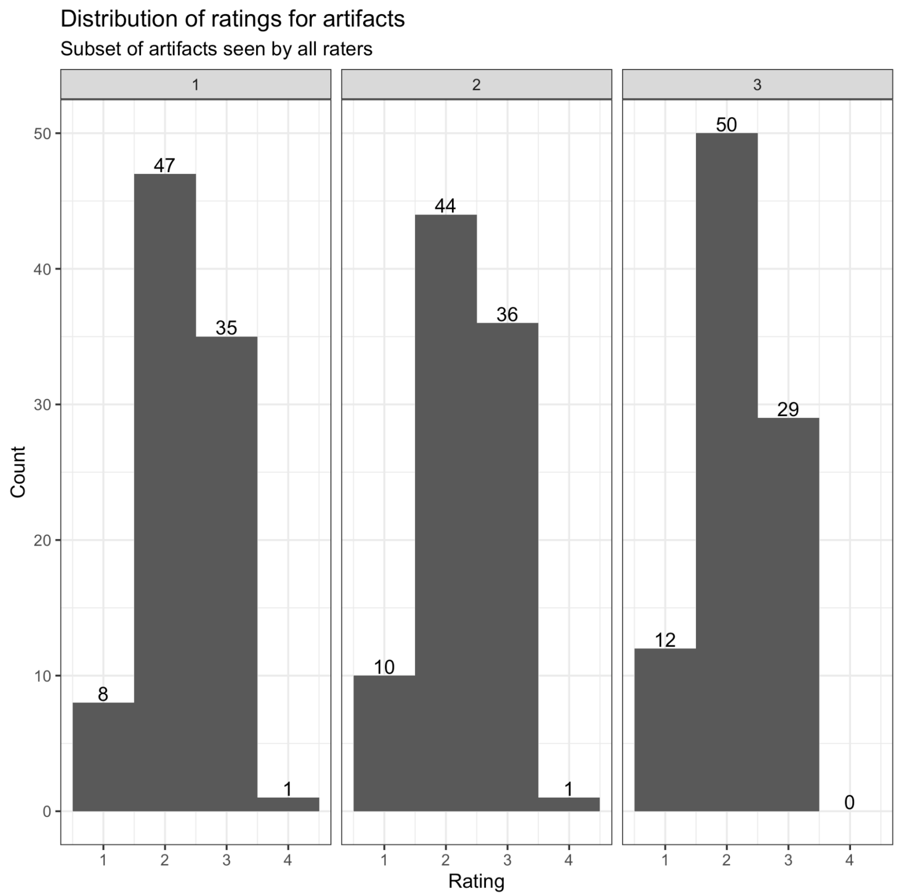


Figure 3: Distribution of ratings between raters for artifacts seen by all raters.

(2) Intraclass Correlation and Percent Exact Agreement

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.The formula to calculate ICC is as follows:

$$ICC = \frac{\boldsymbol{\tau}^2}{\boldsymbol{\tau}^2 + \sigma^2}.$$

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. We then used the summary outputs for each model to find $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}$. The fitted models and actual ICC calculations can be found in Appendix 2. In table 3 below, the final 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. 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 3: 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 4 below. From the table we can see 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%.

Rubric	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 4: 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 3. 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.

(3) Variable Selection

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 5. 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 5: 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 3 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) 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 in figure 5 for the subset data and in Appendix 4 for the full data. Looking at figure 5, we can see that the Fall semester has significantly more artifacts than the spring semester. **This may be the reason Semester becomes a significant fixed effect in the final model and the Spring decreases ratings when compared to the Fall.** For both semesters a rating of 2 is given the most and a rating of 4 is given the least. The same conclusions apply for the histogram of the full dataset. The distribution for ratings by Male and Female for the subset data can be seen in figure 6 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.

We also wanted to visualize the differences between ratings by each rater and each of the rubrics (Figure 4). 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.

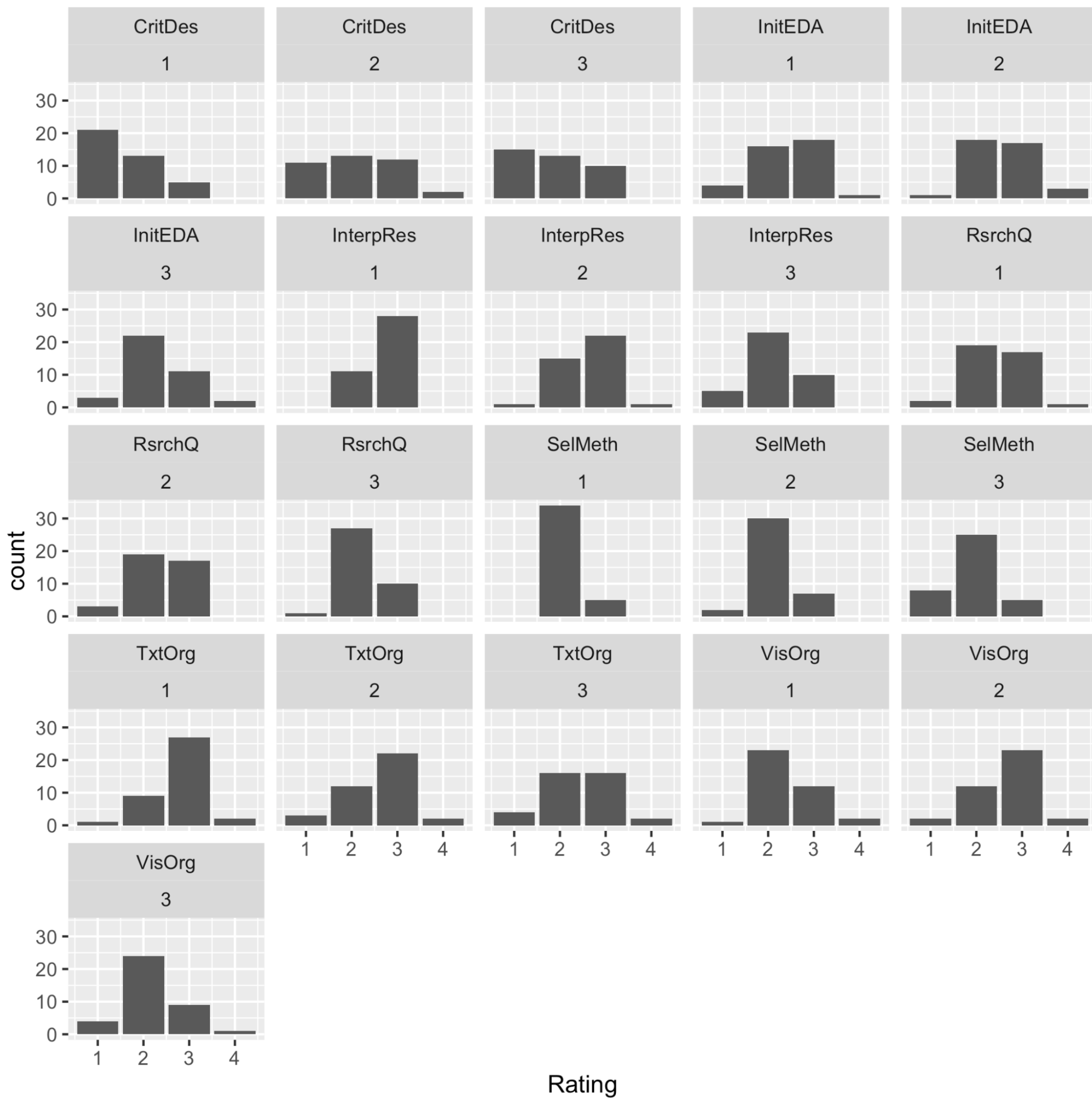


Figure 4: Distribution of ratings by rubric and rater.

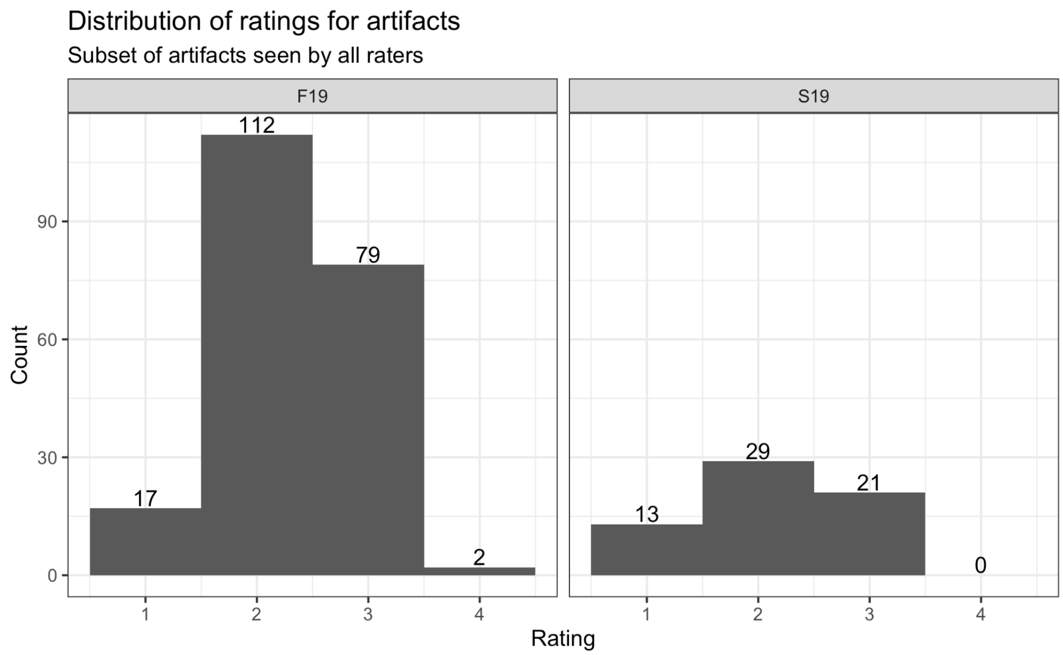


Figure 5: Distribution of ratings by rubric and rater.

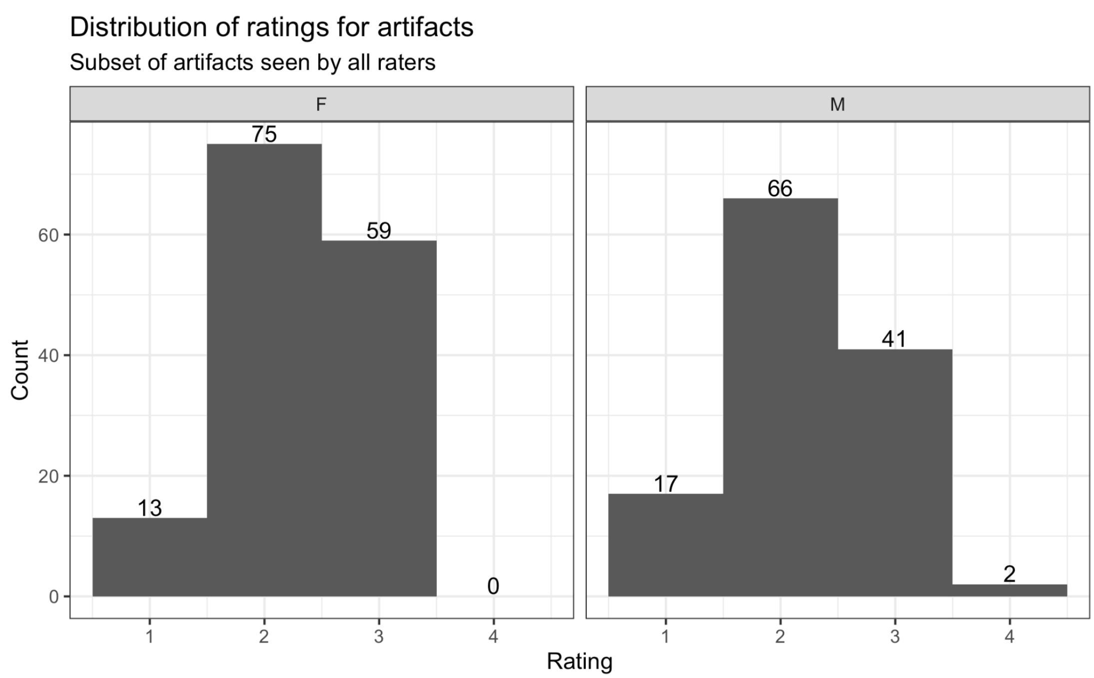


Figure 6: Distribution of ratings by rubric and rater.

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. 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.

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.

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.

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. This 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.

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.

References

Kutner, M.H., Nachsheim, C.J., Neter, J. & Li, W. (2005) Applied Linear Statistical Models, Fifth Edition. NY. McGraw-Hill/Irwin.

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.

https://www.cmu.edu/dietrich/gened/previous-curriculum/index.html (Not sure how to cite websites)

Technical Appendix

Alana Willis

11/13/2021

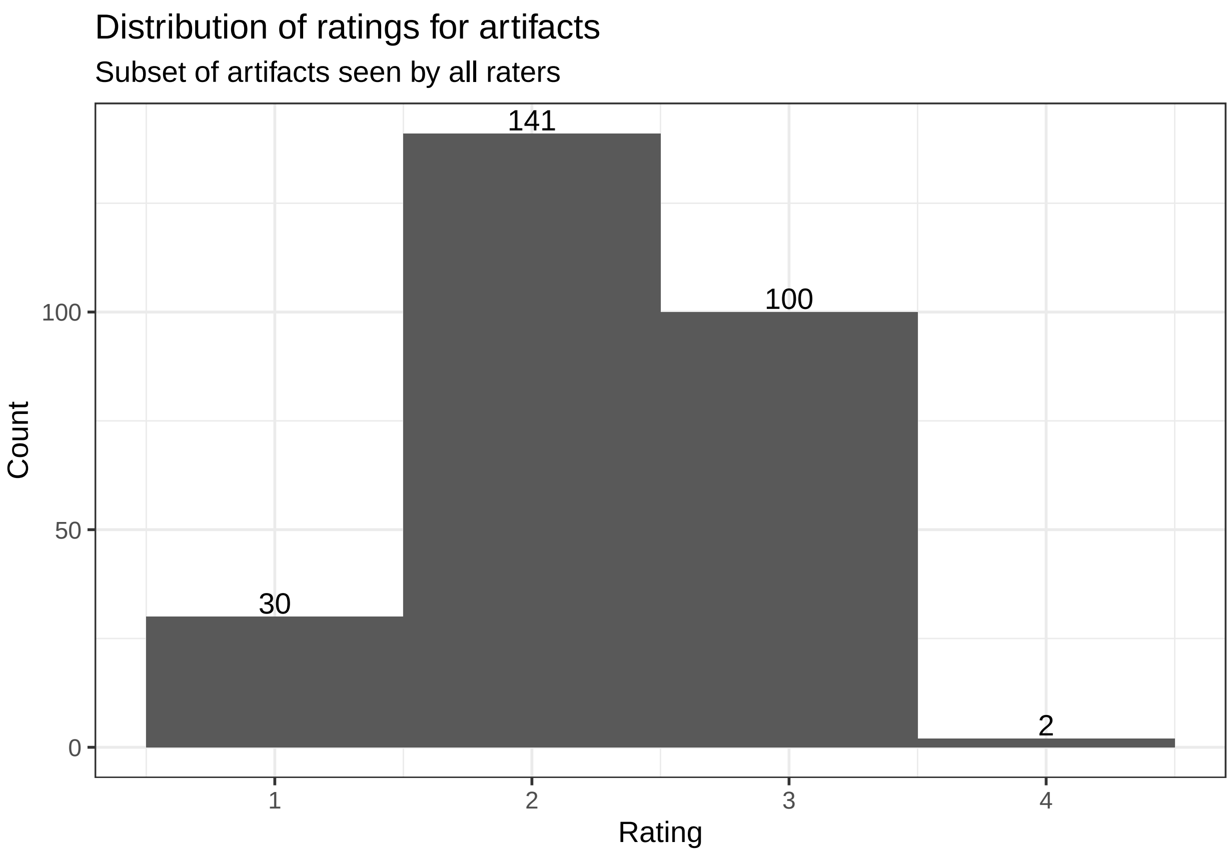
Contents


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

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

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

Appendix 1. Initial Data/Library Imports & Data Exploration

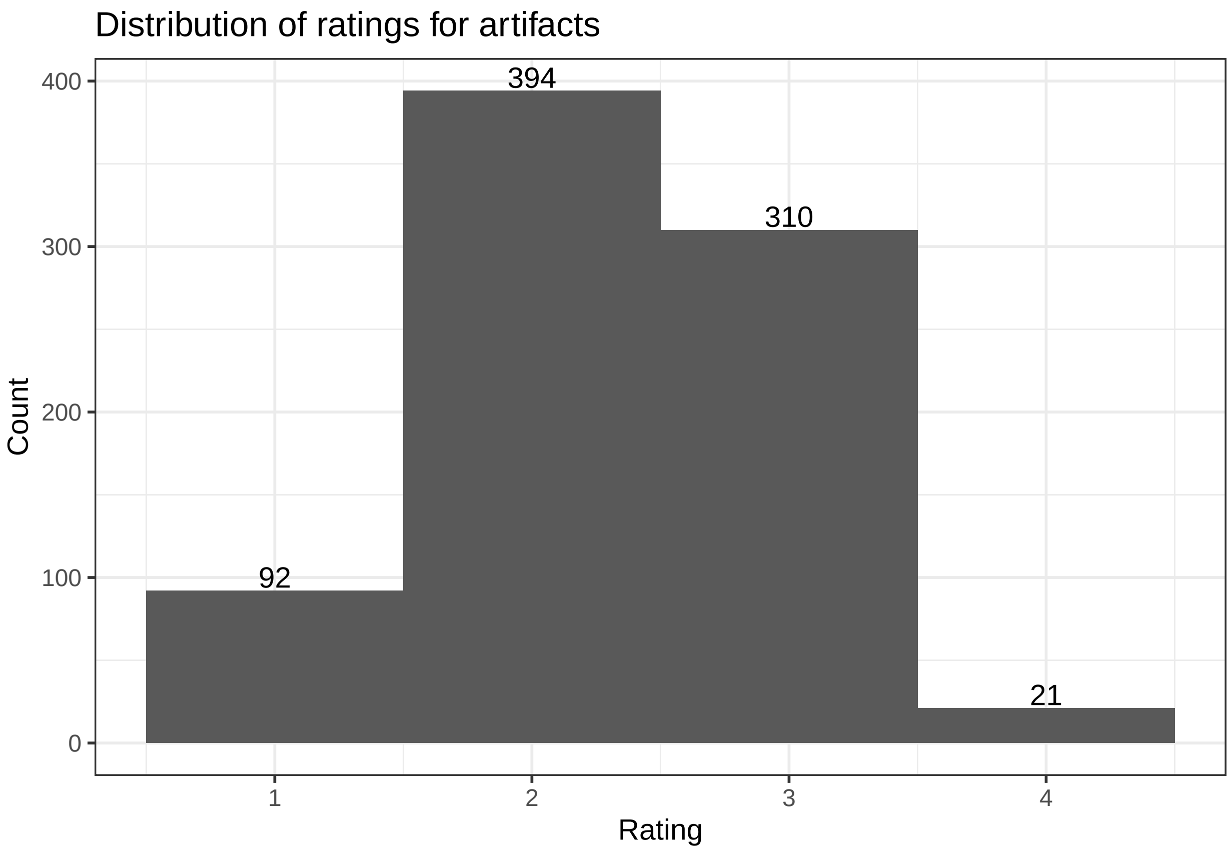
To begin, we loaded the original full data set and created a modified data set that is a subset of just the artifacts seen by all three raters. After this we began EDA.

```
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")

tall <- tall.full %>%
  filter(Repeated == "1") %>%
  dplyr::select(-c(X))

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
  theme_bw()
```

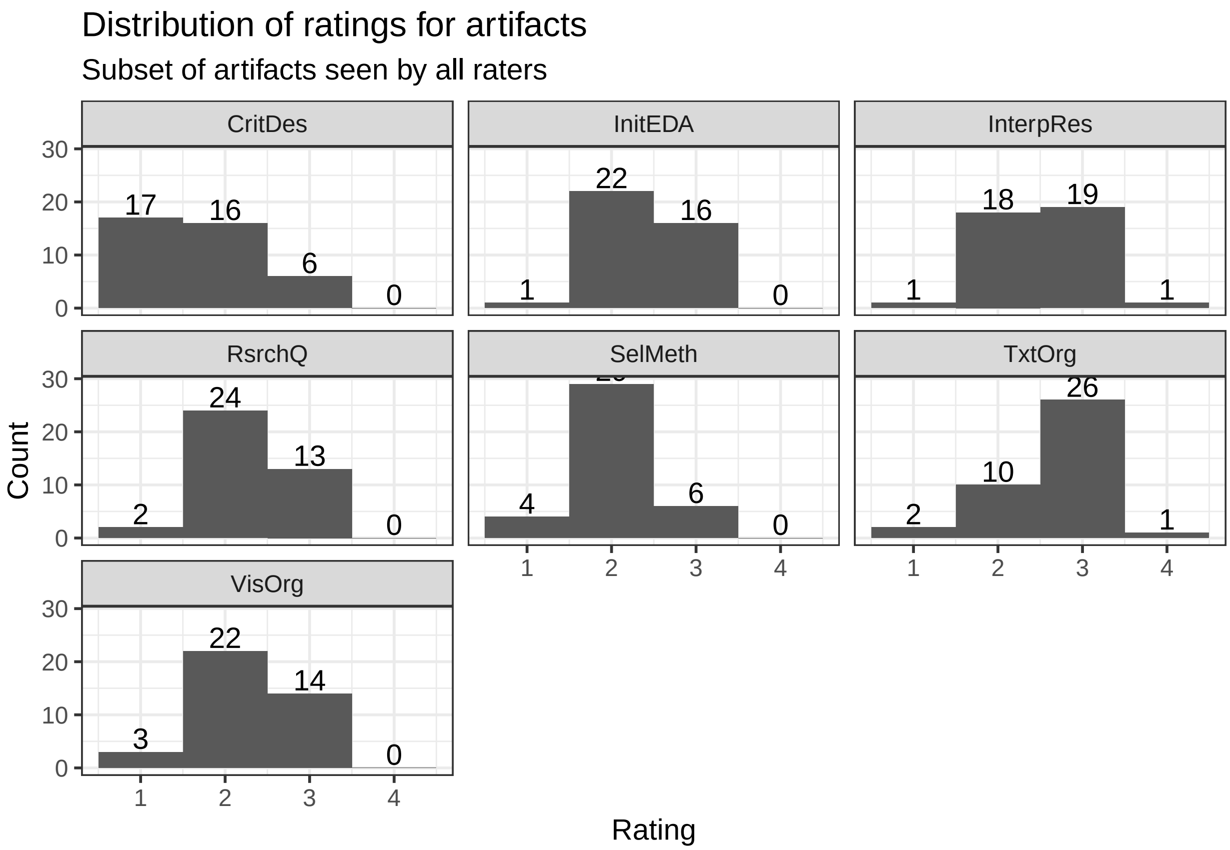



```
tall.full %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw()
```

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

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

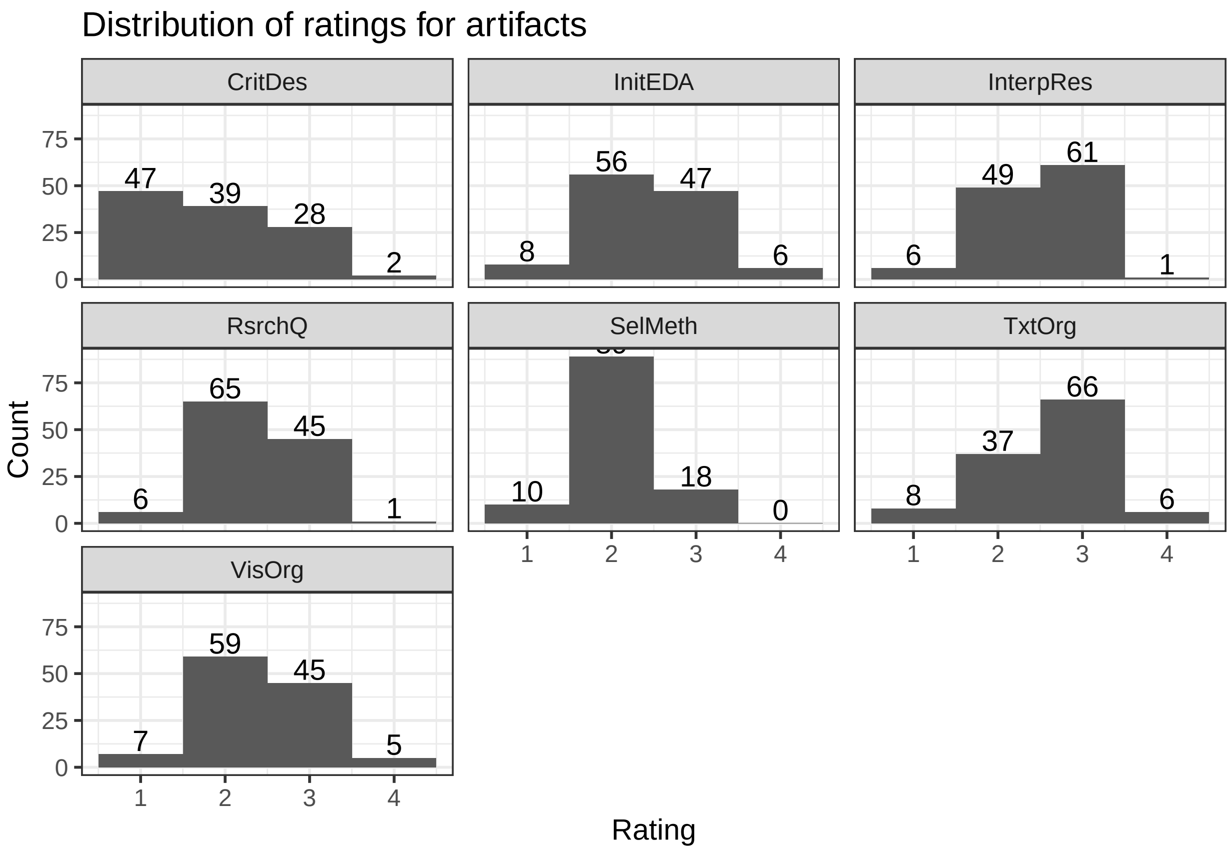



```
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts", subtitle = "Subset of
  theme_bw() +
  facet_wrap(~ Rubric)
```

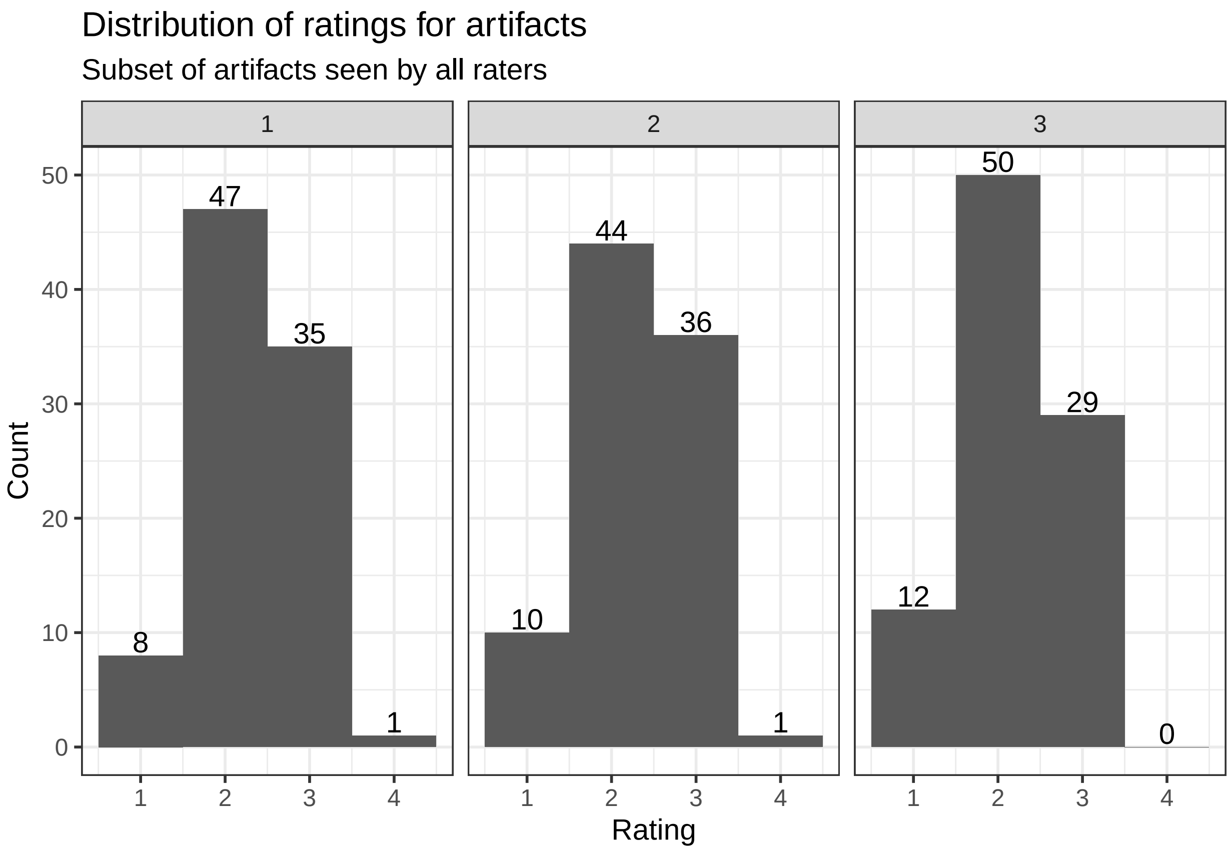



```r
tall.full %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts",) +
  theme_bw() +
  facet_wrap(~ Rubric)
```

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

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

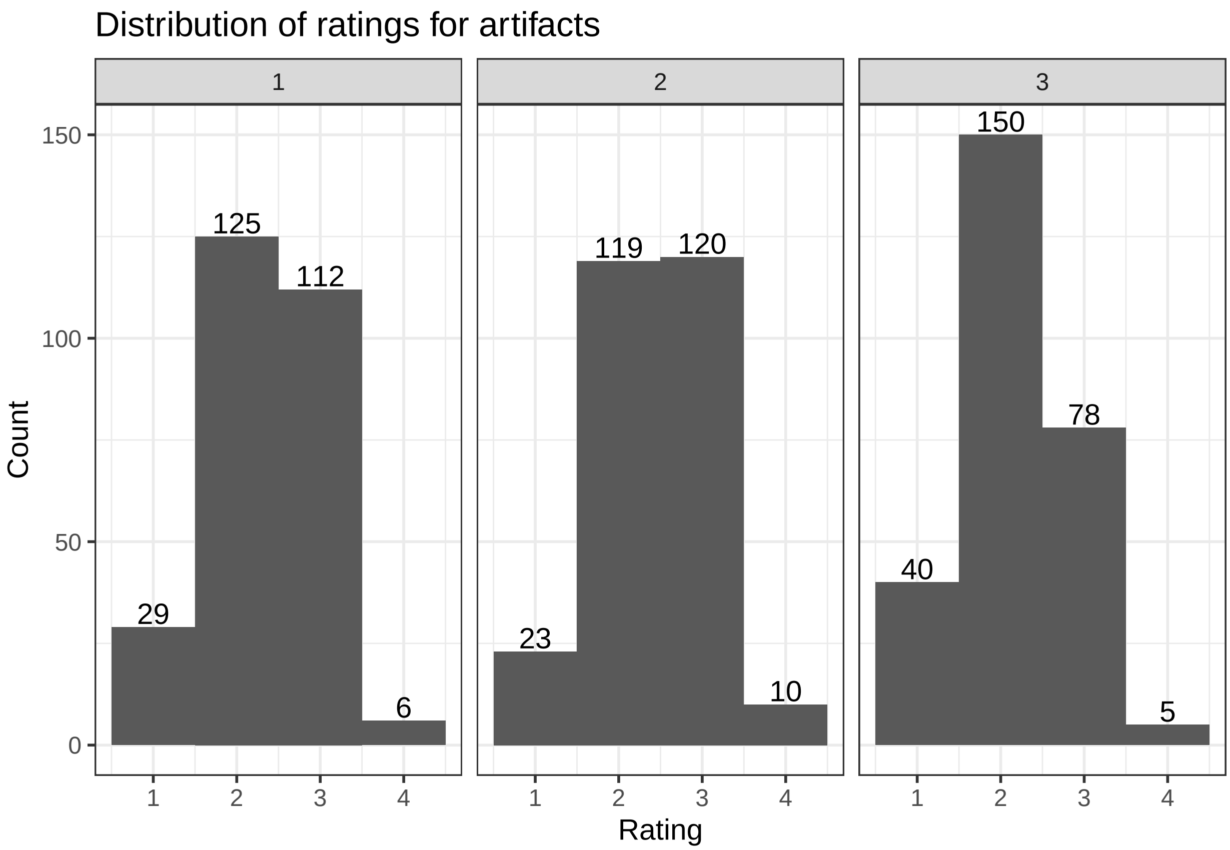



```
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts", subtitle = "Subset of
  theme_bw() +
  facet_wrap(~ Rater)
```




```
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
```

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

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

```
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
```

```
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==
#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
```

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

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

```
CritDes.raters12 <- data.frame(r1=repeated$CritDes[repeated$Rater==1],r2=repeated$CritDes[repeated$Rate
#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
```

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

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

```
InitEDA.raters12 <- data.frame(r1=repeated$InitEDA[repeated$Rater==1],r2=repeated$InitEDA[repeated$Rate
#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
```

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

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

```
SelMeth.raters12 <- data.frame(r1=repeated$SelMeth[repeated$Rater==1],r2=repeated$SelMeth[repeated$Rate
#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$InterpRes[repeatec
#with(InterpRes.raters12,table(r1,r2))
r1 <- factor(InterpRes.raters12$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==
#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
```

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

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

```r
TxtOrg.raters12 <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],r2=repeated$TxtOrg[repeated$Rater==
#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
#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$CritDes[repeated$Rate
#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$InitEDA[repeated$Rate
#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$SelMeth[repeated$Rate
#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$InterpRes[repeate
#with(InterpRes.raters13,table(r1,r3))
r1 <- factor(InterpRes.raters13$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==
#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==
#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==
#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$CritDes[repeated$Rate
#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
```

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

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

```r
InitEDA.raters23 <- data.frame(r2=repeated$InitEDA[repeated$Rater==2],r3=repeated$InitEDA[repeated$Rate
#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
```

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

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

```r
SelMeth.raters23 <- data.frame(r2=repeated$SelMeth[repeated$Rater==2],r3=repeated$SelMeth[repeated$Rate
#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
```

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

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

```r
InterpRes.raters23 <- data.frame(r2=repeated$InterpRes[repeated$Rater==2],r3=repeated$InterpRes[repeate
#with(InterpRes.raters23,table(r2,r3))
r2 <- factor(InterpRes.raters23$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
```

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

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

```
VisOrg.raters23 <- data.frame(r2=repeated$VisOrg[repeated$Rater==2],r3=repeated$VisOrg[repeated$Rater==
#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
```

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

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

```
TxtOrg.raters23 <- data.frame(r2=repeated$TxtOrg[repeated$Rater==2],r3=repeated$TxtOrg[repeated$Rater==
#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.

```r
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
```

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

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

```r
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
```

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

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

```
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
```

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

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

```
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
```

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

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

```
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.

```r
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: 209.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2694 -0.5285 -0.3736  0.9743  2.4770
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07276  0.2697
##  Residual             0.27825  0.5275
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.35169    0.05794   40.59
```

```r
RsrchQ.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=RsrchQ
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 -0.5407 -0.2972  0.8804  2.4434
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
```

```
##  Artifact (Intercept) 0.06873  0.2622
##  Residual             0.28470  0.5336
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.43996    0.11983  20.362
## as.factor(Rater)2  2.35641    0.12150  19.395
## as.factor(Rater)3  2.25527    0.12361  18.245
## SemesterS19        0.08684    0.13225   0.657
## SexM              -0.05845    0.12200  -0.479
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.423
## as.fctr(R)3  0.426  0.435
## SemesterS19 -0.466 -0.448 -0.451
## SexM        -0.519 -0.552 -0.556  0.288
```

```
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        3 211.21 219.47 -102.61   205.21
## RsrchQ.lmer.allfixed    7 216.08 235.35 -101.04   202.08 3.1369  4     0.5352
```

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

```
## 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       3 211.21 219.47 -102.61   205.21
## RsrchQ.lmer.fixed.3    6 214.32 230.84 -101.16   202.32 2.8911  3     0.4087
```

```
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:
```

	df	AIC	BIC
RsrchQ.lmer.full	3	215.0797	223.3404
RsrchQ.lmer.allfixed	7	229.4872	248.7624
RsrchQ.lmer.fixed.3	6	225.3442	241.8658
RsrchQ.lmer.fixed.2	5	221.7319	235.4999

```
## 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       3 211.21 219.47 -102.61   205.21
## RsrchQ.lmer.fixed.2    5 213.04 226.81 -101.52   203.04 2.1767  2     0.3368
```

```
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(R
  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: 274.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.00615 -0.60064  0.02999  0.67713  2.06614
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4909   0.7006
##  Residual             0.2412   0.4911
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.89533    0.08889   21.32
```

```
CritDes.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=CritD
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  0.60486  1.59203
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4462   0.6680
##  Residual             0.2477   0.4977
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  1.65749    0.16523  10.031
## as.factor(Rater)2  2.07748    0.16767  12.390
## as.factor(Rater)3  1.85659    0.16925  10.970
## SemesterS19       -0.03638    0.19723  -0.184
## SexM               0.09744    0.18186   0.536
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.594
## as.fctr(R)3  0.599  0.602
## SemesterS19 -0.511 -0.484 -0.498
## SexM        -0.576 -0.601 -0.601  0.294
```

```
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        3 277.68 285.91 -135.84   271.68
## CritDes.lmer.allfixed    7 277.19 296.41 -131.60   263.19 8.4849  4    0.07535
##
## CritDes.lmer.full
## CritDes.lmer.allfixed .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
CritDes.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact), data=CritDes.rati
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: 272.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.56370 -0.51160 -0.04644  0.62729  1.60451
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4410   0.6641
##  Residual             0.2475   0.4975
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  1.70806    0.13459   12.69
## as.factor(Rater)2  2.13182    0.13354   15.96
## as.factor(Rater)3  1.91116    0.13480   14.18
## SemesterS19       -0.06754    0.18775   -0.36
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.378
## as.fctr(R)3  0.384  0.374
## SemesterS19 -0.437 -0.401 -0.420
```

```
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       3 277.68 285.91 -135.84   271.68
## CritDes.lmer.fixed.3    6 275.49 291.96 -131.75   263.49 8.186  3    0.04232 *
## ---
## 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: 271
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

	df	AIC	BIC
CritDes.lmer.full	3	280.6857	288.9205
CritDes.lmer.allfixed	7	287.6376	306.8521
CritDes.lmer.fixed.3	6	284.3508	300.8204
CritDes.lmer.fixed.2	5	280.9688	294.6934

```
## -1.55495 -0.50027 -0.08228  0.64663  1.60935
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4349   0.6595
##  Residual             0.2473   0.4972
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   1.6863     0.1207   13.98
## as.factor(Rater)2   2.1129     0.1219   17.34
## as.factor(Rater)3   1.8908     0.1219   15.51
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.244
## as.fctr(R)3 0.244  0.246
```

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

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

```
## Data: CritDes.ratings.full1
## Models:
## 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       3 277.68 285.91 -135.84   271.68
## CritDes.lmer.fixed.2    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## ---
## 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),BI
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```r
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: 239
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8889 -0.3391 -0.1427  0.4276  1.6035
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3651   0.6042
##  Residual             0.1655   0.4068
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44226    0.07537    32.4
```

```r
InitEDA.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=InitED
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
## -2.0735 -0.3543 -0.1151  0.3813  1.5146
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3878   0.6227
##  Residual             0.1548   0.3934
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.482936   0.144446  17.189
## as.factor(Rater)2  2.495302   0.145899  17.103
## as.factor(Rater)3  2.289614   0.147598  15.513
## SemesterS19       -0.008809   0.173677  -0.051
## SexM               0.047533   0.161547   0.294
##
## Correlation of Fixed Effects:
```

```
##              a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.640
## as.fctr(R)3  0.641  0.645
## SemesterS19 -0.525 -0.513 -0.514
## SexM        -0.591 -0.611 -0.612  0.306
```

```
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         3 241.64 249.90 -117.82   235.64
## InitEDA.lmer.allfixed     7 246.08 265.36 -116.04   232.08 3.5603  4     0.4688
```

```
InitEDA.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact), data=InitEDA.ratin
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: 242.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0776 -0.3724 -0.1169  0.3798  1.5081
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3814   0.6175
##  Residual             0.1553   0.3941
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.5072     0.1161  21.601
## as.factor(Rater)2   2.5221     0.1150  21.923
## as.factor(Rater)3   2.3164     0.1162  19.931
## SemesterS19        -0.0244     0.1644  -0.148
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.433
## as.fctr(R)3  0.435  0.430
## SemesterS19 -0.448 -0.432 -0.434
```

```
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       3 241.64 249.9 -117.82   235.64
## InitEDA.lmer.fixed.3    6 244.17 260.7 -116.09   232.17 3.4699  3     0.3247
```

```
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: 240.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0739 -0.3708 -0.1231  0.3730  1.5104
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3748   0.6122
##  Residual             0.1557   0.3946
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.4985     0.1034   24.17
## as.factor(Rater)2   2.5154     0.1034   24.33
## as.factor(Rater)3   2.3091     0.1043   22.13
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.294
## as.fctr(R)3 0.295  0.295
```

```
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       3 241.64 249.90 -117.82   235.64
## InitEDA.lmer.fixed.2    5 242.20 255.96 -116.10   232.20 3.4472  2     0.1784
```

	df	AIC	BIC
InitEDA.lmer.full	3	244.9824	253.2431
InitEDA.lmer.allfixed	7	258.0518	277.3269
InitEDA.lmer.fixed.3	6	254.3250	270.8465
InitEDA.lmer.fixed.2	5	250.5683	264.3362

```
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),BI
  kbl(booktabs=T,col.names=c("df","AIC","BIC")) %>%
  kable_minimal(full_width=F)
```

```
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: 153.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.21774 -0.09510 -0.09510 -0.04934  2.11906
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1060   0.3256
##  Residual             0.1227   0.3502
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.0621     0.0485   42.52
```

```
SelMeth.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=SelMe
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: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.18753    0.08956  24.425
## as.factor(Rater)2  2.15942    0.09071  23.806
## as.factor(Rater)3  1.96482    0.09212  21.329
## SemesterS19       -0.31955    0.10246  -3.119
## SexM               0.12159    0.09502   1.280
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.513
## as.fctr(R)3  0.515  0.522
## SemesterS19 -0.494 -0.477 -0.480
## SexM        -0.550 -0.578 -0.581  0.299
```

```
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        3 155.37 163.63 -74.687   149.37
## SelMeth.lmer.allfixed    7 142.35 161.63 -64.178   128.35 21.018  4   0.000314
##
## 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.rati
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: 143.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.0480 -0.3923 -0.0551  0.2674  2.5827
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08973  0.2996
##  Residual             0.10842  0.3293
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.25037    0.07503  29.992
## as.factor(Rater)2  2.22653    0.07424  29.991
## as.factor(Rater)3  2.03316    0.07521  27.033
## SemesterS19       -0.35860    0.09796  -3.661
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.285
## as.fctr(R)3  0.287  0.280
## SemesterS19 -0.413 -0.391 -0.394
```

```
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     6 142.05 158.58 -65.027   130.05
## SelMeth.lmer.allfixed    7 142.35 161.63 -64.178   128.35 1.6988  1     0.1924
```

```
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: 153.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8747 -0.2101 -0.1821  0.1742  2.5827
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1074   0.3277
##  Residual             0.1145   0.3383
## Number of obs: 116, groups:  Artifact, 90
##
```

```
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.13775    0.07184   29.76
## as.factor(Rater)2  2.11942    0.07184   29.50
## as.factor(Rater)3  1.92505    0.07266   26.49
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.160
## as.fctr(R)3 0.162  0.162
```

```
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     5 153.09 166.85 -71.543   143.09
## SelMeth.lmer.allfixed    7 142.35 161.63 -64.178   128.35 14.73  2   0.000633
##
## SelMeth.lmer.fixed.2
## SelMeth.lmer.allfixed ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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       3 155.37 163.63 -74.687   149.37
## SelMeth.lmer.fixed.3    6 142.05 158.58 -65.027   130.05 19.32  3  0.0002348
##
## SelMeth.lmer.full
## SelMeth.lmer.fixed.3 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

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

	df	AIC	BIC
SelMeth.lmer.full	3	159.5943	167.8551
SelMeth.lmer.allfixed	7	158.7925	178.0677
SelMeth.lmer.fixed.3	6	155.5577	172.0792
SelMeth.lmer.fixed.2	5	163.4636	177.2315

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

```
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)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings.full1
##
## REML criterion at convergence: 216.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1331 -0.6911  0.5205  0.7508  2.6562
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08286  0.2879
##  Residual             0.29166  0.5401
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    2.479      0.060   41.32
```

```
InterpRes.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=Int
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 -0.7521  0.2727  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: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.75633    0.11388  24.203
## as.factor(Rater)2  2.63881    0.11546  22.854
## as.factor(Rater)3  2.19402    0.11747  18.678
## SemesterS19       -0.09219    0.12586  -0.733
## SexM              -0.05770    0.11614  -0.497
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.426
## as.fctr(R)3  0.429  0.438
## SemesterS19 -0.467 -0.449 -0.452
## SexM        -0.520 -0.553 -0.557  0.289
```

```
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        3 218.53 226.79 -106.263   212.53
## InterpRes.lmer.allfixed    7 204.01 223.28  -95.003   190.01 22.519  4
##                         Pr(>Chisq)
## InterpRes.lmer.full
## InterpRes.lmer.allfixed  0.0001579 ***
## ---
## 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.
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: 201.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5443 -0.7754  0.2680  0.6220  2.6305
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06525  0.2554
##  Residual             0.25154  0.5015
```

```
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1  2.72685    0.09696  28.125
## as.factor(Rater)2  2.60709    0.09589  27.188
## as.factor(Rater)3  2.16161    0.09726  22.225
## SemesterS19       -0.07419    0.12013  -0.618
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.195
## as.fctr(R)3  0.197  0.189
## SemesterS19 -0.388 -0.363 -0.366
```

```
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     6 202.27 218.79 -95.134   190.27
## InterpRes.lmer.allfixed    7 204.01 223.28 -95.003   190.01 0.2615  1
##                         Pr(>Chisq)
## InterpRes.lmer.fixed.3
## InterpRes.lmer.allfixed     0.6091
```

```
InterpRes.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact), data=InterpRes.ratings.full
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: 199.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5317 -0.7627  0.2635  0.6614  2.6535
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06224  0.2495
##  Residual             0.25250  0.5025
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1  2.70421    0.08912   30.34
```

	df	AIC	BIC
InterpRes.lmer.full	3	222.3211	230.5819
InterpRes.lmer.allfixed	7	217.9267	237.2019
InterpRes.lmer.fixed.3	6	213.7029	230.2244
InterpRes.lmer.fixed.2	5	209.6794	223.4474

```
## as.factor(Rater)2  2.58574    0.08912   29.01
## as.factor(Rater)3  2.13918    0.09027   23.70
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.061
## as.fctr(R)3 0.062  0.062
```

```
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    5 200.66 214.43 -95.331   190.66
## InterpRes.lmer.fixed.3    6 202.27 218.79 -95.134   190.27 0.3936  1     0.5304
```

```
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
  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: 224.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5890 -0.3728 -0.1605  0.4763  1.6335
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3106   0.5573
##  Residual             0.1589   0.3987
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.43901    0.07114   34.29
```

```
VisOrg.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=VisOrg
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  0.4059  1.8478
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2797   0.5289
##  Residual             0.1494   0.3865
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.5445     0.1324  19.219
## as.factor(Rater)2   2.8145     0.1322  21.295
## as.factor(Rater)3   2.4487     0.1339  18.291
## SemesterS19        -0.2458     0.1546  -1.590
## SexM               -0.1992     0.1439  -1.384
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.609
## as.fctr(R)3  0.611  0.614
## SemesterS19 -0.524 -0.509 -0.511
## SexM        -0.587 -0.607 -0.609  0.313
```

```
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        3 227.21 235.44 -110.60   221.21
## VisOrg.lmer.allfixed    7 221.36 240.57 -103.68   207.36 13.848  4   0.007795
##
## VisOrg.lmer.full
## VisOrg.lmer.allfixed **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
VisOrg.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact), data=VisOrg.ratings
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: 220.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5232 -0.3853 -0.1405  0.4201  1.8217
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2872   0.5359
##  Residual             0.1475   0.3840
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.4361     0.1076  22.632
## as.factor(Rater)2   2.7032     0.1055  25.631
## as.factor(Rater)3   2.3377     0.1066  21.925
## SemesterS19        -0.1787     0.1479  -1.208
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.400
## as.fctr(R)3  0.402  0.394
## SemesterS19 -0.444 -0.425 -0.427
```

```
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       3 227.21 235.44 -110.60   221.21
## VisOrg.lmer.fixed.3    6 221.33 237.80 -104.66   209.33 11.881  3   0.007802 **
## ---
## 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: 219.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5004 -0.3365 -0.2483  0.3841  1.8552
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2907   0.5392
##  Residual             0.1467   0.3830
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.37794    0.09658   24.62
## as.factor(Rater)2  2.64891    0.09564   27.70
## as.factor(Rater)3  2.28355    0.09658   23.64
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2
## as.fctr(R)2 0.263
## as.fctr(R)3 0.265  0.263
```

```r
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       3 227.21 235.44 -110.60   221.21
## VisOrg.lmer.fixed.2    5 220.82 234.54 -105.41   210.82 10.392  2   0.005539 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

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

	df	AIC	BIC
VisOrg.lmer.full	3	230.6621	238.8969
VisOrg.lmer.allfixed	7	234.2432	253.4577
VisOrg.lmer.fixed.3	6	232.1098	248.5793
VisOrg.lmer.fixed.2	5	229.5832	243.3079

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

```
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: 247.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3557 -0.7550  0.3834  0.5302  2.4132
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09371  0.3061
##  Residual             0.39573  0.6291
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.58745    0.06821   37.93
```

```
TxtOrg.lmer.allfixed <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact), data=TxtOrg
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  0.2551  0.5680  2.3900
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06472  0.2544
```

```
## Residual             0.41037  0.6406
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.86126    0.13842  20.671
## as.factor(Rater)2  2.68224    0.14036  19.110
## as.factor(Rater)3  2.52450    0.14283  17.675
## SemesterS19       -0.20904    0.15157  -1.379
## SexM              -0.08794    0.13960  -0.630
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 SmsS19
## as.fctr(R)2  0.403
## as.fctr(R)3  0.406  0.416
## SemesterS19 -0.459 -0.441 -0.444
## SexM        -0.512 -0.546 -0.550  0.285
```

```
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        3 250.00 258.26 -122.00   244.00
## TxtOrg.lmer.allfixed    7 251.06 270.34 -118.53   237.06 6.934  4     0.1394
```

```
TxtOrg.lmer.fixed.3 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact), data=TxtOrg.ratings
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: 247.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4299 -0.6362  0.2467  0.5657  2.2838
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07216  0.2686
##  Residual             0.40104  0.6333
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1   2.8157     0.1187  23.730
```

```
## as.factor(Rater)2   2.6332     0.1174  22.438
## as.factor(Rater)3   2.4754     0.1190  20.793
## SemesterS19        -0.1819     0.1455  -1.251
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.178
## as.fctr(R)3  0.179  0.171
## SemesterS19 -0.383 -0.357 -0.361
```

```
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       3 250.00 258.26 -122.00   244.00
## TxtOrg.lmer.fixed.3    6 249.49 266.01 -118.74   237.49 6.5101  3    0.08926 .
## ---
## 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: 246.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3524 -0.5901  0.3213  0.5582  2.1044
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07484  0.2736
##  Residual             0.40093  0.6332
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1   2.7586     0.1099   25.10
## as.factor(Rater)2   2.5806     0.1099   23.48
## as.factor(Rater)3   2.4218     0.1113   21.76
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.049
## as.fctr(R)3 0.050  0.050
```

	df	AIC	BIC
TxtOrg.lmer.full	3	253.5354	261.7962
TxtOrg.lmer.allfixed	7	263.0165	282.2916
TxtOrg.lmer.fixed.3	6	259.3060	275.8275
TxtOrg.lmer.fixed.2	5	256.8503	270.6183

```
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       3 250.00 258.26 -122.00   244.00
## TxtOrg.lmer.fixed.2    5 249.09 262.86 -119.55   239.09 4.9021  2     0.0862 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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(T
  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.ratin
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: 143.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0480 -0.3923 -0.0551  0.2674  2.5827
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08973  0.2996
##  Residual             0.10842  0.3293
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.25037    0.07503  29.992
## as.factor(Rater)2  2.22653    0.07424  29.991
## as.factor(Rater)3  2.03316    0.07521  27.033
## SemesterS19       -0.35860    0.09796  -3.661
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.285
## as.fctr(R)3  0.287  0.280
## SemesterS19 -0.413 -0.391 -0.394
```

```
SelMeth.lmer.int <-lmer(Rating ~ -1 + as.factor(Rater)*Semester - Semester + (1|Artifact), data=SelMeth
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: 144.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1978 -0.3605 -0.1393  0.1645  2.4202
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08661  0.2943
##  Residual             0.10999  0.3317
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                               Estimate Std. Error t value
## as.factor(Rater)1              2.21374    0.08179  27.066
## as.factor(Rater)2              2.21370    0.08047  27.511
## as.factor(Rater)3              2.08256    0.08179  25.462
## as.factor(Rater)1:SemesterS19 -0.23481    0.14843  -1.582
## as.factor(Rater)2:SemesterS19 -0.31119    0.15198  -2.048
## as.factor(Rater)3:SemesterS19 -0.53832    0.15268  -3.526
##
## Correlation of Fixed Effects:
##             as.(R)1 as.(R)2 as.(R)3 a.(R)1: a.(R)2:
## as.fctr(R)2  0.155
## as.fctr(R)3  0.157   0.155
## a.(R)1:SS19 -0.551  -0.086  -0.087
## a.(R)2:SS19 -0.082  -0.529  -0.082   0.124
## a.(R)3:SS19 -0.084  -0.083  -0.536   0.125   0.126
```

```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    6 142.05 158.58 -65.027   130.05
## SelMeth.lmer.int        8 143.46 165.49 -63.731   127.46 2.592  2     0.2736
```

```r
# SelMeth.lmer.ran1 <-lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact) + (Semester|Artifac
# summary(SelMeth.lmer.ran1)
# SelMeth.lmer.ran2 <- lmer(Rating ~ -1 + as.factor(Rater) + Semester + (1|Artifact) + (as.factor(Rater
# summary(SelMeth.lmer.ran2)


InterpRes.lmer.fixed.2 <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact), data=InterpRes.ratings.ful
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: 199.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5317 -0.7627  0.2635  0.6614  2.6535
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06224  0.2495
##  Residual             0.25250  0.5025
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.70421    0.08912   30.34
## as.factor(Rater)2  2.58574    0.08912   29.01
## as.factor(Rater)3  2.13918    0.09027   23.70
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.061
## as.fctr(R)3 0.062  0.062
```

```r
# InterpRes.lmer.ran <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact)+ (as.factor(Rater)|Artifact),
# 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: 219.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5004 -0.3365 -0.2483  0.3841  1.8552
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2907   0.5392
##  Residual             0.1467   0.3830
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##                   Estimate Std. Error t value
## as.factor(Rater)1  2.37794    0.09658   24.62
## as.factor(Rater)2  2.64891    0.09564   27.70
## as.factor(Rater)3  2.28355    0.09658   23.64
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.263
## as.fctr(R)3 0.265  0.263
```

```
# VisOrg.lmer.ran <-lmer(Rating ~ -1 + as.factor(Rater) + (1|Artifact) + (as.factor(Rater)|Artifact), d
# 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.

```
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| = 0.00711231 (tol = 0.002, component 1)
```

```
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
##   1527.0   1668.0   -733.5   1467.0      780
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0158 -0.4919 -0.0810  0.5258  3.7868
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.63877  0.7992
##           RubricInitEDA   0.38193  0.6180    0.25
##           RubricInterpRes 0.25558  0.5055   -0.01  0.78
##           RubricRsrchQ    0.17263  0.4155    0.38  0.50  0.74
##           RubricSelMeth   0.09483  0.3079    0.56  0.36  0.40  0.25
##           RubricTxtOrg    0.40339  0.6351    0.02  0.69  0.80  0.64  0.23
##           RubricVisOrg    0.31788  0.5638    0.17  0.78  0.77  0.59  0.28  0.79
##  Residual                 0.19449  0.4410
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.23153    0.03989   55.94
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00711231 (tol = 0.002, component 1)
```

```
all.fixed.lmer <- lmer(Rating ~ -1 + as.factor(Rater) + Repeated + Semester + Sex + Rubric + (0 + Rubri
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
##   1467.5   1660.1   -692.7   1385.5      769
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1241 -0.5076 -0.0206  0.5323  3.8103
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.54163  0.7360
##           RubricInitEDA   0.34303  0.5857   0.46
##           RubricInterpRes 0.16723  0.4089   0.22 0.75
##           RubricRsrchQ    0.16225  0.4028   0.58 0.43 0.70
##           RubricSelMeth   0.06233  0.2497   0.38 0.59 0.73 0.38
##           RubricTxtOrg    0.25464  0.5046   0.33 0.61 0.70 0.55 0.66
##           RubricVisOrg    0.24917  0.4992   0.34 0.73 0.67 0.50 0.39 0.75
##  Residual                 0.18746  0.4330
## Number of obs: 810, groups:  Artifact, 90
##
```

```
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1  2.014358   0.107640  18.714
## as.factor(Rater)2  2.016736   0.108019  18.670
## as.factor(Rater)3  1.838073   0.108457  16.947
## Repeated          -0.073455   0.095519  -0.769
## SemesterS19       -0.175051   0.085553  -2.046
## SexM               0.009812   0.079126   0.124
## RubricInitEDA      0.547092   0.095162   5.749
## RubricInterpRes    0.587063   0.100309   5.853
## RubricRsrchQ       0.460912   0.087046   5.295
## RubricSelMeth      0.164924   0.093845   1.757
## RubricTxtOrg       0.692978   0.098934   7.004
## RubricVisOrg       0.530063   0.098580   5.377
##
## Correlation of Fixed Effects:
##              a.(R)1 a.(R)2 a.(R)3 Repetd SmsS19 SexM   RbIEDA RbrcIR RbrcRQ
## as.fctr(R)2  0.873
## as.fctr(R)3  0.873  0.873
## Repeated    -0.152 -0.151 -0.152
## SemesterS19 -0.356 -0.351 -0.353  0.079
## SexM        -0.392 -0.404 -0.407  0.009  0.301
## RubrcIntEDA -0.556 -0.555 -0.552  0.008 -0.001  0.000
## RbrcIntrpRs -0.666 -0.664 -0.661 -0.010 -0.001  0.000  0.734
## RubrcRsrchQ -0.632 -0.630 -0.627 -0.040 -0.001  0.000  0.585  0.756
## RubricSlMth -0.695 -0.693 -0.690 -0.089 -0.001  0.000  0.658  0.776  0.689
## RubrcTxtOrg -0.616 -0.614 -0.611  0.005 -0.001  0.000  0.674  0.752  0.682
## RubricVsOrg -0.612 -0.610 -0.608 -0.022 -0.002 -0.001  0.715  0.745  0.668
##              RbrcSM RbrcTO
## as.fctr(R)2
## as.fctr(R)3
## Repeated
## SemesterS19
## SexM
## RubrcIntEDA
## RbrcIntrpRs
## RubrcRsrchQ
## RubricSlMth
## RubrcTxtOrg  0.725
## RubricVsOrg  0.680  0.751
```

```
fixed.lmer <-lmer(Rating ~ -1 + as.factor(Rater) + Rubric + Semester + (0 + Rubric | Artifact), data=tal
```

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

```
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
```

```
##   1464.0   1647.2   -693.0   1386.0      771
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1337 -0.5145 -0.0183  0.5363  3.7906
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.54592  0.7389
##           RubricInitEDA   0.34377  0.5863   0.46
##           RubricInterpRes 0.16422  0.4052   0.23 0.75
##           RubricRsrchQ    0.16310  0.4039   0.59 0.43 0.70
##           RubricSelMeth   0.06213  0.2492   0.39 0.60 0.73 0.39
##           RubricTxtOrg    0.25008  0.5001   0.33 0.61 0.69 0.55 0.65
##           RubricVisOrg    0.25308  0.5031   0.35 0.73 0.68 0.51 0.41 0.75
##  Residual                 0.18777  0.4333
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                    Estimate Std. Error t value
## as.factor(Rater)1  2.00847    0.09801  20.493
## as.factor(Rater)2  2.00894    0.09780  20.540
## as.factor(Rater)3  1.83067    0.09804  18.673
## RubricInitEDA      0.54754    0.09517   5.753
## RubricInterpRes    0.58642    0.10028   5.848
## RubricRsrchQ       0.45840    0.08693   5.273
## RubricSelMeth      0.15902    0.09328   1.705
## RubricTxtOrg       0.69315    0.09896   7.004
## RubricVisOrg       0.52876    0.09852   5.367
## SemesterS19       -0.17274    0.08155  -2.118
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3 RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO
## as.fctr(R)2  0.846
## as.fctr(R)3  0.845  0.845
## RubrcIntEDA -0.612 -0.613 -0.611
## RbrcIntrpRs -0.737 -0.739 -0.737  0.734
## RubrcRsrchQ -0.703 -0.705 -0.703  0.586  0.756
## RubricSlMth -0.783 -0.785 -0.783  0.662  0.779  0.689
## RubrcTxtOrg -0.681 -0.682 -0.680  0.674  0.751  0.682  0.728
## RubricVsOrg -0.676 -0.678 -0.676  0.716  0.745  0.667  0.681  0.751
## SemesterS19 -0.263 -0.254 -0.256 -0.002  0.000  0.002  0.006 -0.001  0.000
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
int.lmer <- lmer(Rating ~ -1 + as.factor(Rater)*Rubric + as.factor(Rater)*Semester + Semester*Rubric +
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00349687 (tol = 0.002, component 1)
```

```
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
##   1458.3   1735.4   -670.1   1340.3      751
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9464 -0.5411 -0.0584  0.5131  3.6754
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.49178  0.7013
##           RubricInitEDA   0.34250  0.5852   0.44
##           RubricInterpRes 0.14347  0.3788   0.34 0.81
##           RubricRsrchQ    0.15983  0.3998   0.64 0.42 0.71
##           RubricSelMeth   0.06244  0.2499   0.43 0.63 0.78 0.45
##           RubricTxtOrg    0.25395  0.5039   0.41 0.63 0.66 0.56 0.63
##           RubricVisOrg    0.24751  0.4975   0.33 0.71 0.67 0.53 0.36 0.76
##  Residual                 0.17774  0.4216
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                                  Estimate Std. Error t value
## as.factor(Rater)1                1.714010   0.128399  13.349
## as.factor(Rater)2                2.088602   0.127442  16.389
## as.factor(Rater)3                1.955412   0.128540  15.212
## RubricInitEDA                    0.731381   0.149165   4.903
## RubricInterpRes                  1.012599   0.147053   6.886
## RubricRsrchQ                     0.687518   0.133653   5.144
## RubricSelMeth                    0.511363   0.141466   3.615
## RubricTxtOrg                     1.081986   0.147785   7.321
## RubricVisOrg                     0.714826   0.152447   4.689
## SemesterS19                     -0.019019   0.197665  -0.096
## as.factor(Rater)2:RubricInitEDA   -0.304240   0.169557  -1.794
## as.factor(Rater)3:RubricInitEDA   -0.293487   0.169818  -1.728
## as.factor(Rater)2:RubricInterpRes -0.531325   0.167258  -3.177
## as.factor(Rater)3:RubricInterpRes -0.748025   0.167603  -4.463
## as.factor(Rater)2:RubricRsrchQ    -0.489494   0.157923  -3.100
## as.factor(Rater)3:RubricRsrchQ    -0.362179   0.158129  -2.290
## as.factor(Rater)2:RubricSelMeth   -0.399701   0.161519  -2.475
## as.factor(Rater)3:RubricSelMeth   -0.411556   0.161826  -2.543
## as.factor(Rater)2:RubricTxtOrg    -0.583080   0.168655  -3.457
## as.factor(Rater)3:RubricTxtOrg    -0.484414   0.168952  -2.867
## as.factor(Rater)2:RubricVisOrg    -0.143662   0.171561  -0.837
## as.factor(Rater)3:RubricVisOrg    -0.330594   0.171882  -1.923
## as.factor(Rater)2:SemesterS19     -0.043923   0.121550  -0.361
## as.factor(Rater)3:SemesterS19     -0.118364   0.121659  -0.973
## RubricInitEDA:SemesterS19          0.045942   0.203885   0.225
## RubricInterpRes:SemesterS19       -0.002161   0.201063  -0.011
## RubricRsrchQ:SemesterS19           0.185334   0.178009   1.041
## RubricSelMeth:SemesterS19         -0.277866   0.193232  -1.438
## RubricTxtOrg:SemesterS19          -0.108102   0.201451  -0.537
```

```
## RubricVisOrg:SemesterS19          -0.105986   0.208364  -0.509
```

```
##
## 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)
## Model failed to converge with max|grad| = 0.00349687 (tol = 0.002, component 1)
```

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

```
## 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
##   1454.5   1694.1   -676.3   1352.5      759
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9680 -0.5167 -0.0423  0.4870  3.6510
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.49565  0.7040
##           RubricInitEDA   0.34968  0.5913   0.44
##           RubricInterpRes 0.14872  0.3856   0.35 0.81
##           RubricRsrchQ    0.17676  0.4204   0.63 0.44 0.71
##           RubricSelMeth   0.06559  0.2561   0.42 0.59 0.73 0.35
##           RubricTxtOrg    0.25739  0.5073   0.41 0.63 0.67 0.54 0.63
##           RubricVisOrg    0.25090  0.5009   0.34 0.71 0.67 0.51 0.37 0.77
##  Residual                 0.17863  0.4226
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                                    Estimate Std. Error t value
## as.factor(Rater)1                   1.76092    0.11637  15.132
## as.factor(Rater)2                   2.12458    0.11713  18.139
## as.factor(Rater)3                   1.97374    0.11726  16.832
## RubricInitEDA                       0.74568    0.13501   5.523
## RubricInterpRes                     1.01325    0.13303   7.616
## RubricRsrchQ                        0.74800    0.12252   6.105
## RubricSelMeth                       0.42548    0.12871   3.306
## RubricTxtOrg                        1.04785    0.13382   7.830
## RubricVisOrg                        0.68201    0.13773   4.952
## SemesterS19                        -0.17876    0.08099  -2.207
```

```
## as.factor(Rater)2:RubricInitEDA   -0.30785    0.16991  -1.812
## as.factor(Rater)3:RubricInitEDA   -0.29463    0.17023  -1.731
## as.factor(Rater)2:RubricInterpRes -0.53500    0.16749  -3.194
## as.factor(Rater)3:RubricInterpRes -0.75068    0.16789  -4.471
## as.factor(Rater)2:RubricRsrchQ    -0.50012    0.15904  -3.145
## as.factor(Rater)3:RubricRsrchQ    -0.36872    0.15932  -2.314
## as.factor(Rater)2:RubricSelMeth   -0.39403    0.16213  -2.430
## as.factor(Rater)3:RubricSelMeth   -0.41220    0.16249  -2.537
## as.factor(Rater)2:RubricTxtOrg    -0.58193    0.16890  -3.445
## as.factor(Rater)3:RubricTxtOrg    -0.48434    0.16925  -2.862
## as.factor(Rater)2:RubricVisOrg    -0.14213    0.17184  -0.827
## as.factor(Rater)3:RubricVisOrg    -0.33107    0.17222  -1.922
```

```
##
## 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 + a
```

```
## 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
##   1415.9   1683.6   -650.9   1301.9      753
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.10313 -0.47230 -0.03307  0.45501  2.79548
##
## Random effects:
##  Groups     Name              Variance Std.Dev. Corr
##  Artifact   RubricCritDes     0.48901  0.6993
##             RubricInitEDA     0.31273  0.5592    0.32
##             RubricInterpRes   0.09948  0.3154    0.14  0.67
##             RubricRsrchQ      0.17630  0.4199    0.50  0.19  0.53
##             RubricSelMeth     0.03698  0.1923    0.13  0.21  0.36 -0.26
##             RubricTxtOrg      0.24654  0.4965    0.27  0.43  0.36  0.30  0.20
```

```
##             RubricVisOrg      0.22836  0.4779    0.17  0.50  0.44  0.27 -0.18
##  Artifact.1 as.factor(Rater)1 0.01253  0.1119
##             as.factor(Rater)2 0.10956  0.3310   -0.46
##             as.factor(Rater)3 0.09194  0.3032    0.35  0.67
##  Residual                     0.12961  0.3600
##
##
##
##
##
##
##
##   0.53
##
##
##
##
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                                  Estimate Std. Error t value
## as.factor(Rater)1                 1.75880    0.11257  15.624
## as.factor(Rater)2                 2.12276    0.12142  17.483
## as.factor(Rater)3                 1.95373    0.11808  16.546
## RubricInitEDA                     0.73912    0.12833   5.760
## RubricInterpRes                   0.99004    0.12607   7.853
## RubricRsrchQ                      0.72482    0.11641   6.227
## RubricSelMeth                     0.40957    0.12314   3.326
## RubricTxtOrg                      1.01391    0.12840   7.896
## RubricVisOrg                      0.65236    0.13190   4.946
## SemesterS19                      -0.15954    0.07507  -2.125
## as.factor(Rater)2:RubricInitEDA   -0.29935    0.15362  -1.949
## as.factor(Rater)3:RubricInitEDA   -0.29441    0.15388  -1.913
## as.factor(Rater)2:RubricInterpRes -0.51128    0.15101  -3.386
## as.factor(Rater)3:RubricInterpRes -0.71281    0.15116  -4.716
## as.factor(Rater)2:RubricRsrchQ    -0.48594    0.14491  -3.353
## as.factor(Rater)3:RubricRsrchQ    -0.32001    0.14495  -2.208
## as.factor(Rater)2:RubricSelMeth   -0.38454    0.14789  -2.600
## as.factor(Rater)3:RubricSelMeth   -0.38625    0.14718  -2.624
## as.factor(Rater)2:RubricTxtOrg    -0.54890    0.15403  -3.563
## as.factor(Rater)3:RubricTxtOrg    -0.44240    0.15430  -2.867
## as.factor(Rater)2:RubricVisOrg    -0.10229    0.15613  -0.655
## as.factor(Rater)3:RubricVisOrg    -0.27198    0.15636  -1.739
```

```
##
## 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
```

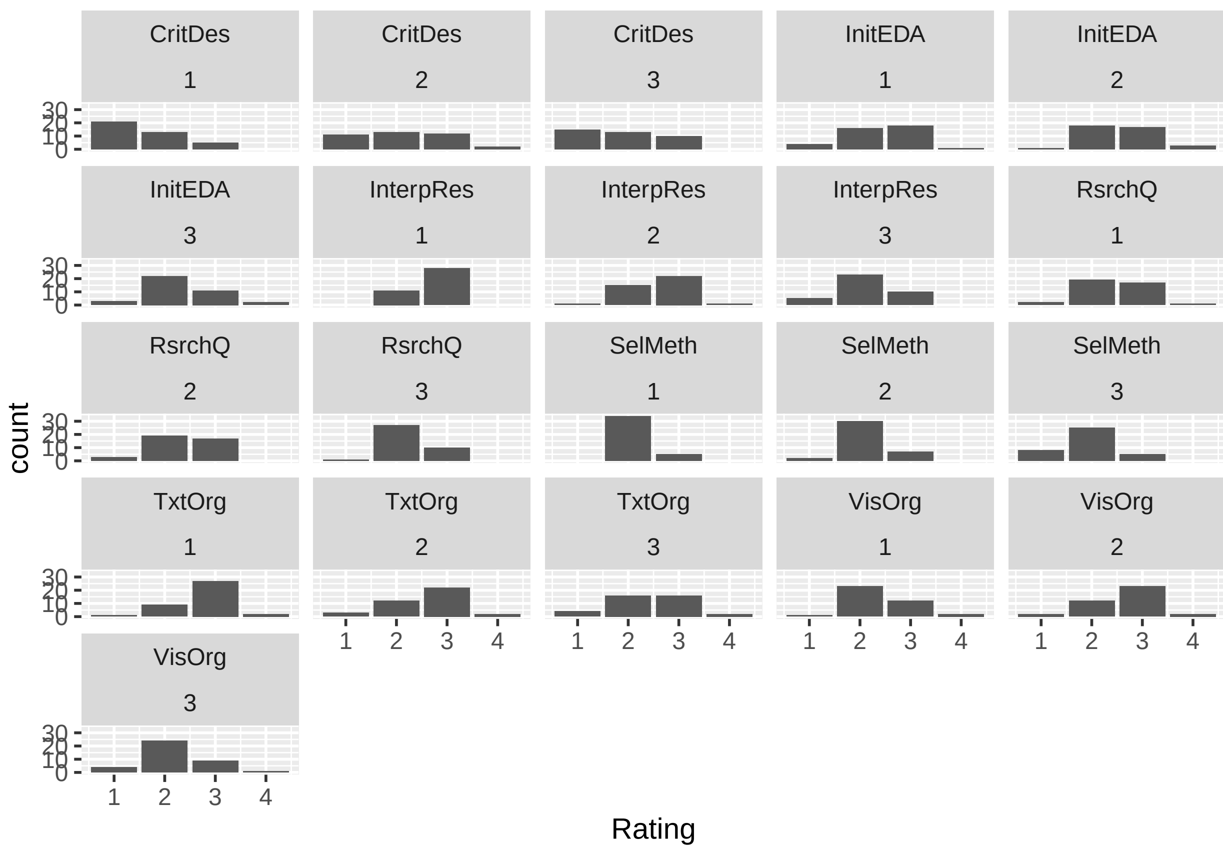
	df	AIC	BIC
null.lmer	30	1527.042	1667.953
all.fixed.lmer	41	1467.486	1660.065
fixed.lmer	39	1464.043	1647.227
int.lmer	59	1458.270	1735.395
int.lmer2	51	1454.517	1694.066
int.ran.lmer	57	1415.870	1683.601

```
data.frame(AIC=AIC(null.lmer,all.fixed.lmer,fixed.lmer,int.lmer, int.lmer2,int.ran.lmer),BIC(null.lmer,a
  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.

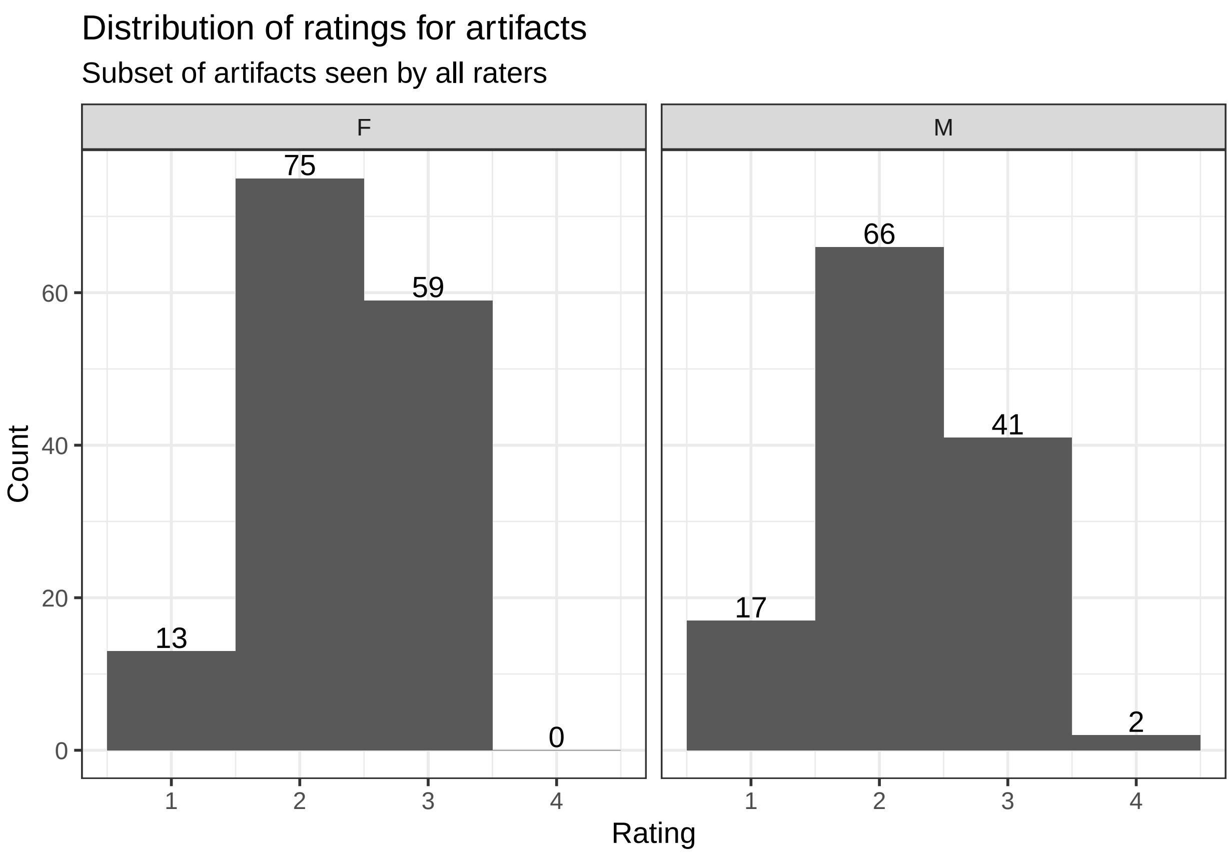
```
ggplot(tall.full.nonmissing, aes(x=Rating)) +
geom_bar() +
facet_wrap( ~ Rubric + Rater)
```



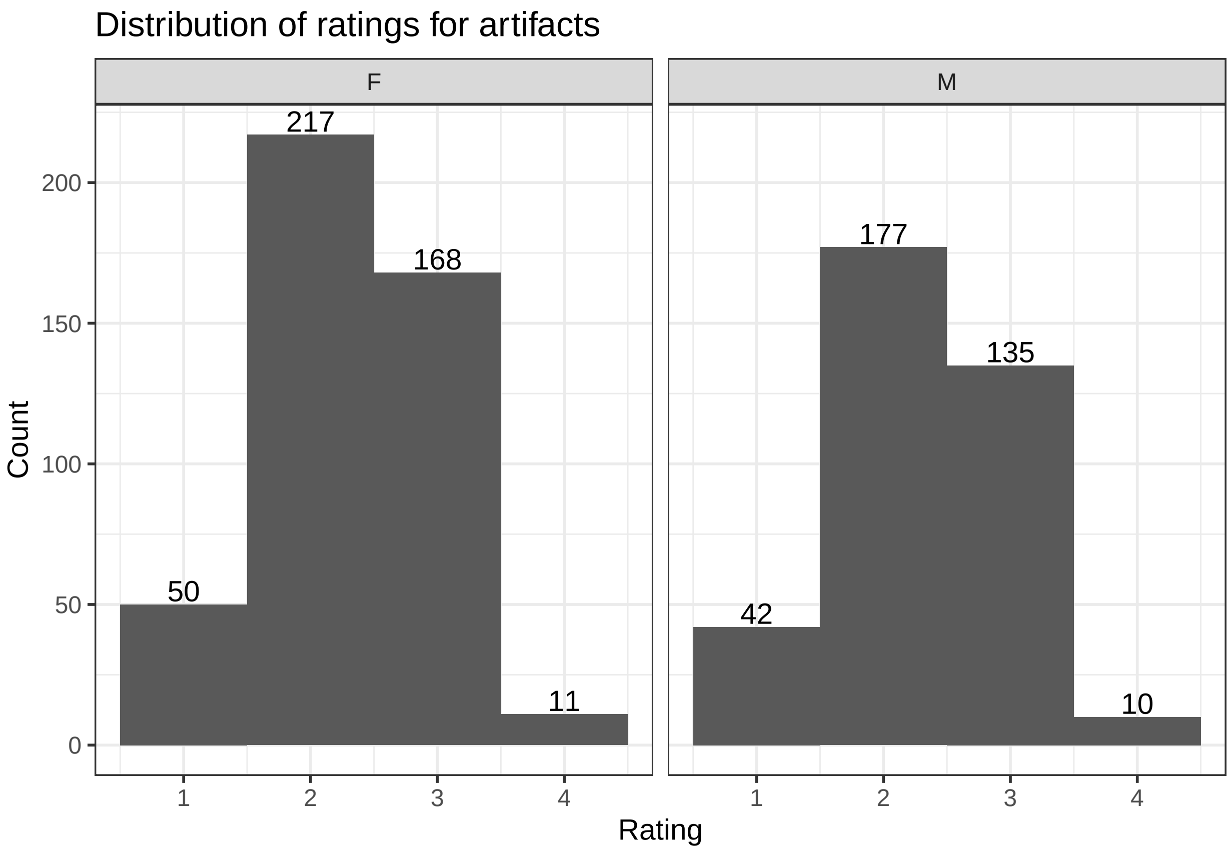
```
ggsave("big.png")
```

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

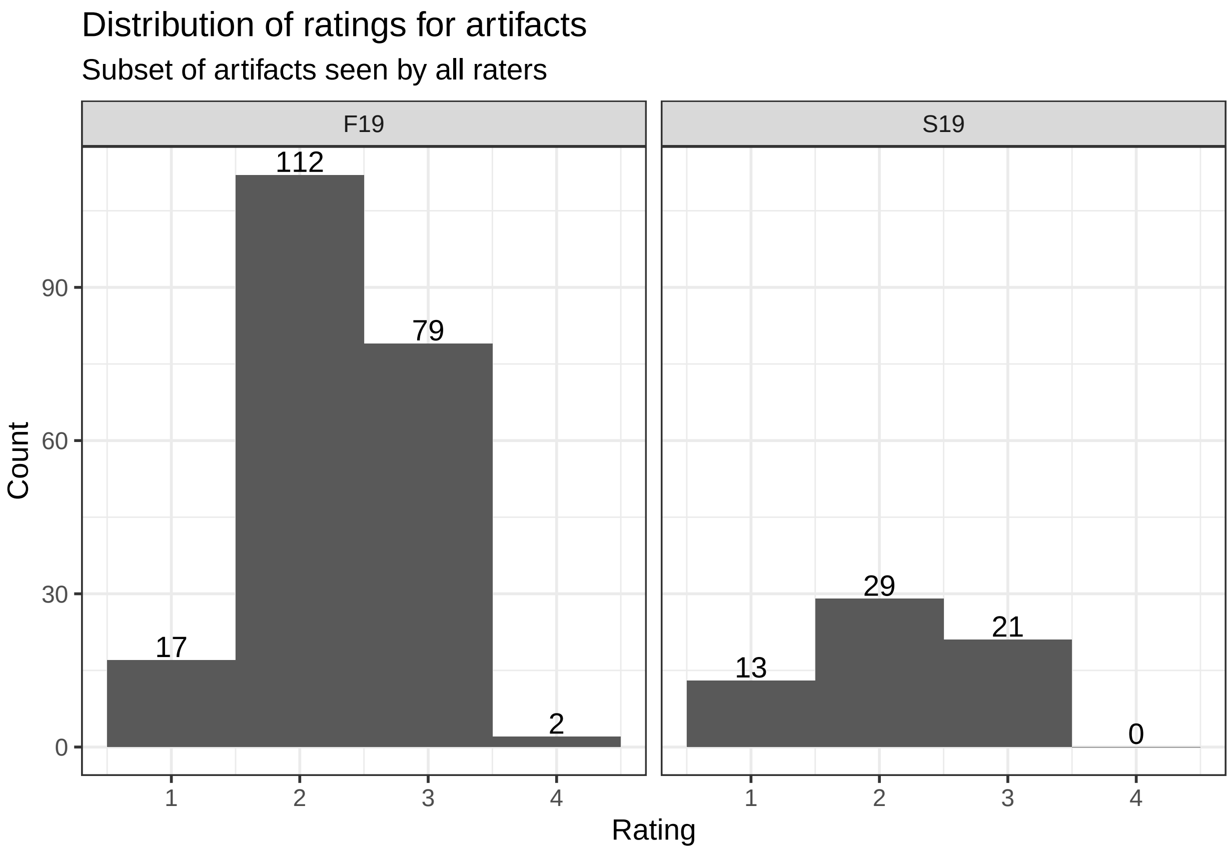
```
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts", subtitle = "Subset of
  theme_bw() +
  facet_wrap(~ Sex)
```




```
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)
```



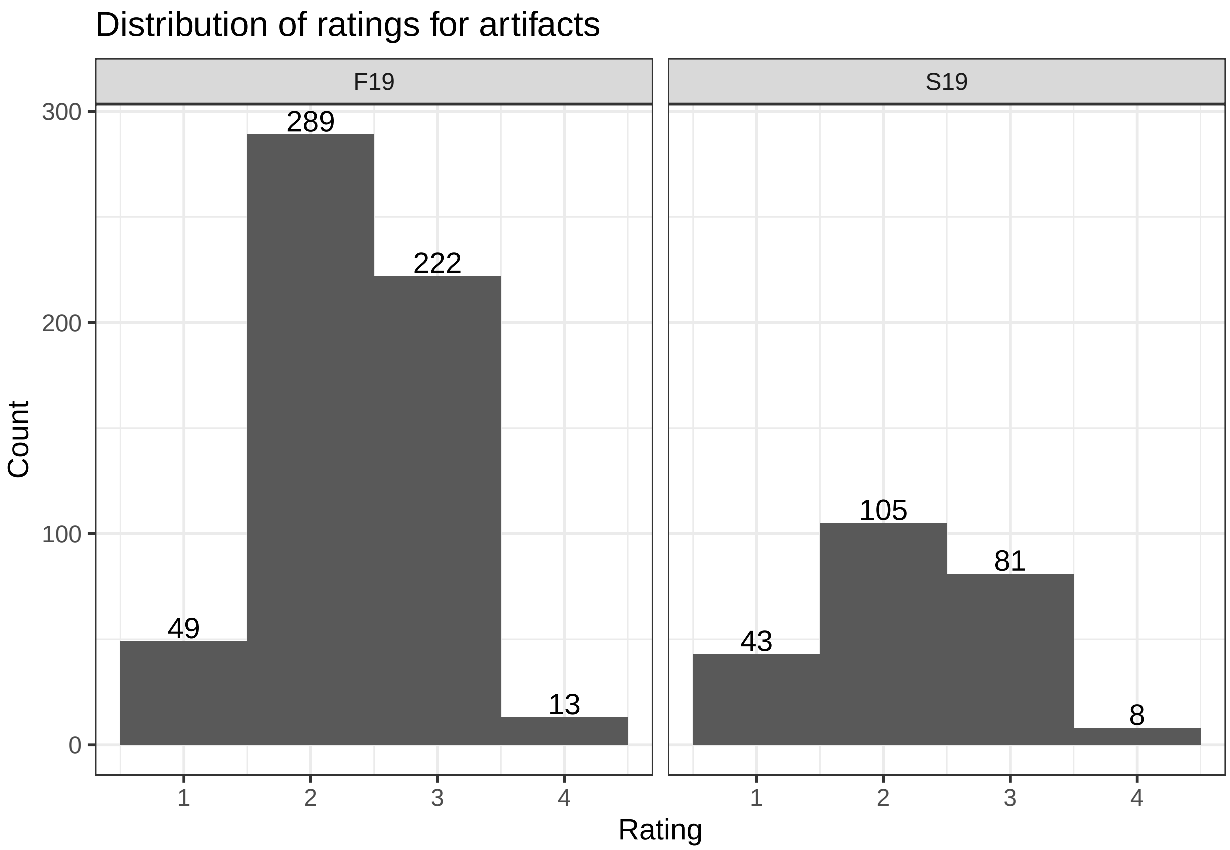

```
tall %>%
  ggplot(aes(x=Rating)) +
  geom_histogram(bins=4) +
  stat_bin(aes(y=..count.., label=..count..), geom="text", vjust=-.15, bins=4) +
  labs(y = "Count", x = "Rating", title= "Distribution of ratings for artifacts", subtitle = "Subset of
  theme_bw() +
  facet_wrap(~ Semester)
```




```
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)
```


Distribution of ratings for artifacts
F19
S19
289
222
49
13
105
81
43
8
300
200
100
0
Count
1
2
3
4
Rating