Investigating the distribution of ratings and relationships with all other factors based on the experiment of rating on General-Education Courses

Yueni Wang |Carnegie Mellon University| yueniw@andrew.cmu.edu

Abstract:

In this paper, we focus on the departmental experiment of anonymous rating on General Education courses. Questions that we want to address are: patterns in distribution of ratings in different rubrics and raters, if raters have largely different opinions within each rubric, the relationships between all factors and ratings through proper model predictions and the reason for skewness in each factors. The data we used comes from three raters from different department grading the courses on seven rubrics. We looked at the indistinguishable pattern of ratings distributions across raters and rubrics and little agreements across some rubrics. The model of of choice should be the one considering random effects. The possible interpretation was that random effect models are more suitable to interpret the ratings scores for each raters across rubrics.

Introduction:

The project is inspired by a new policy that the Dietrich College at Carnegie Mellon University recently conducted. The college is in the process of implementing a new "General Education" program for undergraduates which specifies a set of courses and experiences that all undergraduates must take. To determine whether the new program is successful, the college hopes to rate student work performed in each of the "Gen Ed" courses each year. With the resulting ratings for artifacts on seven rubrics from three raters in three different departments, we are specifically interested in looking into the following questions to provide the dean from Dietrich College as an insight of the efficiency of whole 'Gen Ed' experiment:

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 experiment (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?

4. Why are some factors from rating data being very skewed? How would the pattern affect our result and final model selection?

Data:

The data we are using comes from a designed experiment on General Education feedbacks. 91project papers—referred to as "artifacts"—were randomly sampled from a Fall and Spring section of Fresh-man Statistics. Three raters from three different departments were asked to rate these artifacts on seven rubrics, as shown in Table1. 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: Rubrics for rating Freshman Statistics projects. (Only approved in this experiment)

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 used for all rubrics. (Only approved in this experiment)

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; each of the remaining 78 artifacts were rated by only rater. We are using two csv files in this study, the first csv file with ratings has the following variables shown in table3:

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: Variables in the rating.csv file. (Only approved in this experiment)

The second csv file contains the same data but organized so that each row has one rating listed under the rating column. We want to do some explanatory data analysis on our rating.csv.

X	Rater	Sample	Overlap	Semester	Sex	RsrchQ	CritDes	InitEDA	SelMeth
1	3	1	5	Fall	M	3	3	2	2
2	3	2	7	Fall	F	3	3	3	3
3	3	3	9	Spring	F	2	1	3	2
4	3	4	8	Spring	M	2	2	2	1
5	3	5	NA	Fall	–	3	3	3	3
6	3	6	NA	Fall	M	2	1	2	2

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

Table 4.1 &4.2: The summary statistics for rating.csv.

We look at the summary statistics of ratings. In order to see if there are indistinguishable patterns of ratings distributions across raters and rubrics, we are making subsets according to rubrics and raters and see the distribution and summary statistics on ratings in method section to answer the first research question.

Methods:

Methods for research question one is just the exploratory data analysis on the csv file of ratings. By subsetting the data based on raters and rubrics, we look at the barplots for ratings to see if there's difference in distribution for each subset.

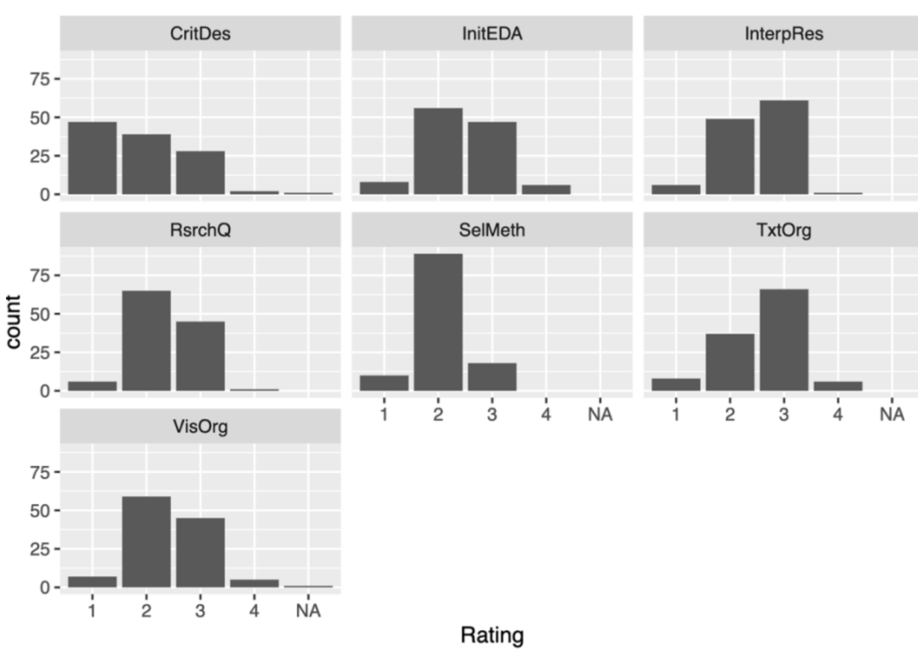
For the second research question, we compare the ICC values across factors after fitting linear mixed-effects models. ICC is the common correlation among the raters' ratings for each artifact. We treat each artifact as a cluster of three ratings and fit the random-intercept model and fit seven random-intercept models, one for each rubric, then calculate the ICCs. Because ICC values could not tell us which raters might be contributing to the disagreement, we are also using the 2-way table to count for the ratings of each pair of raters. The percentage will tell us the exact agreement or disagreement on each rubric.

For the third question, to directly examine interactions with rubric, we are trying to fit linear mixed-effects models along with ANOVA tables and stepwise selections to select the best model. We first try adding fixed effects to Rating ~ (0 + Rubric | Artifact), and then add fixed effects (and possibly interactions) for Rater, Semester, Sex and/or Repeated to the random intercept models and getting rid of the interception term for the tall.csv data, looking at interactions then performing variable selection. Also, we are adding random effects to the starting model to see if models containing random effects would actually be more effective.

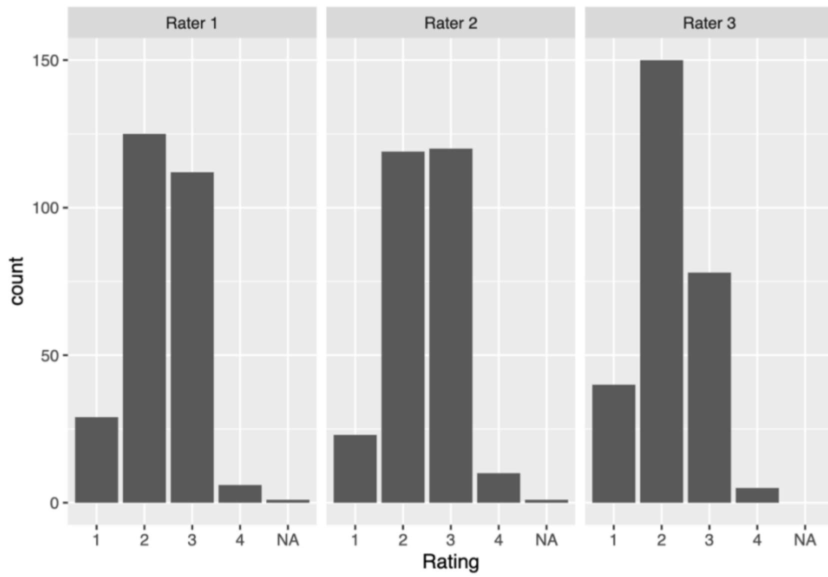
For the fourth question, I am doing exploratory data analysis again and making more data visualizations along with residual plots and correlation plots to see if the extremely skewed or unevenly distributed data has obvious associations with other factors.

Results:

First research question: After Subsetting the data, the resulting bar plots are as follows:



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




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

After dividing the dataset into subsets, we had NA values in our output bar plots and also in our table of counts. It is caused by the missing Sex value and might cause further problem in model fitting. Therefore, we just define the NA value as another Sex value.

Second research question:

```
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.common	a12	a23	a13
CritDes	0.57	0.54	0.69	0.62
InitEDA	0.49	0.69	0.85	0.54
InterpRes	0.23	0.62	0.62	0.54
RsrchQ	0.19	0.38	0.54	0.77
SelMeth	0.52	0.92	0.69	0.62
TxtOrg	0.14	0.69	0.54	0.62
VisOrg	0.59	0.54	0.77	0.77

```
    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
```

```
    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
```

	ICC.alldata	ICC.common	a12	a23	a13
CritDes	0.67	0.57	0.54	0.69	0.62
InitEDA	0.69	0.49	0.69	0.85	0.54
InterpRes	0.22	0.23	0.62	0.62	0.54
RsrchQ	0.21	0.19	0.38	0.54	0.77
SelMeth	0.47	0.52	0.92	0.69	0.62
TxtOrg	0.19	0.14	0.69	0.54	0.62
VisOrg	0.66	0.59	0.54	0.77	0.77

Because ICC values could not tell us which raters might be contributing to the disagreement, we are also using the 2-way table to count for the ratings of each pair of raters. The percentage will tell us the exact agreement or disagreement on each rubric. Here I am including one example of the 2-way tables from one of the seven models and one model summary.

Third research question:

This question is divided into four parts:

(i) Adding fixed effects to the seven rubric-specific models using just the data from the 13 common artifacts that all three raters saw.

```
$CritDes
as.numeric(Rating) ~ (1 | Artifact)

$InitEDA
as.numeric(Rating) ~ (1 | Artifact)

$InterpRes
as.numeric(Rating) ~ (1 | Artifact)

$RsrchQ
as.numeric(Rating) ~ (1 | Artifact)

$SelMeth
as.numeric(Rating) ~ (1 | Artifact)

$TxtOrg
as.numeric(Rating) ~ (1 | Artifact)

$VisOrg
as.numeric(Rating) ~ (1 | Artifact)
```

The starting model was: tmp <- lmer(as.numeric(Rating) ~ -1 + as.factor(Rater) + Semester + Sex + (1|Artifact),data=tall.13[tall.13$Rubric=="RsrchQ",],REML=FALSE)

The final model after adding factors to the single model and using backward elimination to find the model with lowest AIC value. ANOVA table is also used to select best model.

(ii) Trying interactions and new random effects for the seven rubric specific models using all the data

```
$CritDes
as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1

$InitEDA
as.numeric(Rating) ~ (1 | Artifact)

$InterpRes
as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1

$RsrchQ
as.numeric(Rating) ~ (1 | Artifact)

$SelMeth
as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
    1
```

```
$TxtOrg
as.numeric(Rating) ~ (1 | Artifact)

$VisOrg
as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
```

The final model fitted by the above method but using the full dataset is given above.

(iii) trying interactions and new random effects for the seven rubric specific models using all the data

We see there are some differences among the models: For InitEDA, RsrchQ and SelMeth, the models are just the simple random-intercept models. We want to examine each of these 4 models to see if the fixed effects make sense to us and if there are any interactions or additional random effects to consider. After refitting the model and check on the t-statistics, we see the difference across the coefficients for these four factors and decided to keep Rater as an important factor. Adding random effect and perform the model selection again would give us the following model:

```
Linear mixed model fit by REML ['lmerMod']
Formula: as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
    1
   Data: tall.nonmissing[tall.nonmissing$Rubric == "SelMeth", ]

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
```

(iv) Trying to add fixed effects, interactions, and new random effects to the "combined" model Rating ~ 1 + (0 + Rubric | Artifact), using all the data.

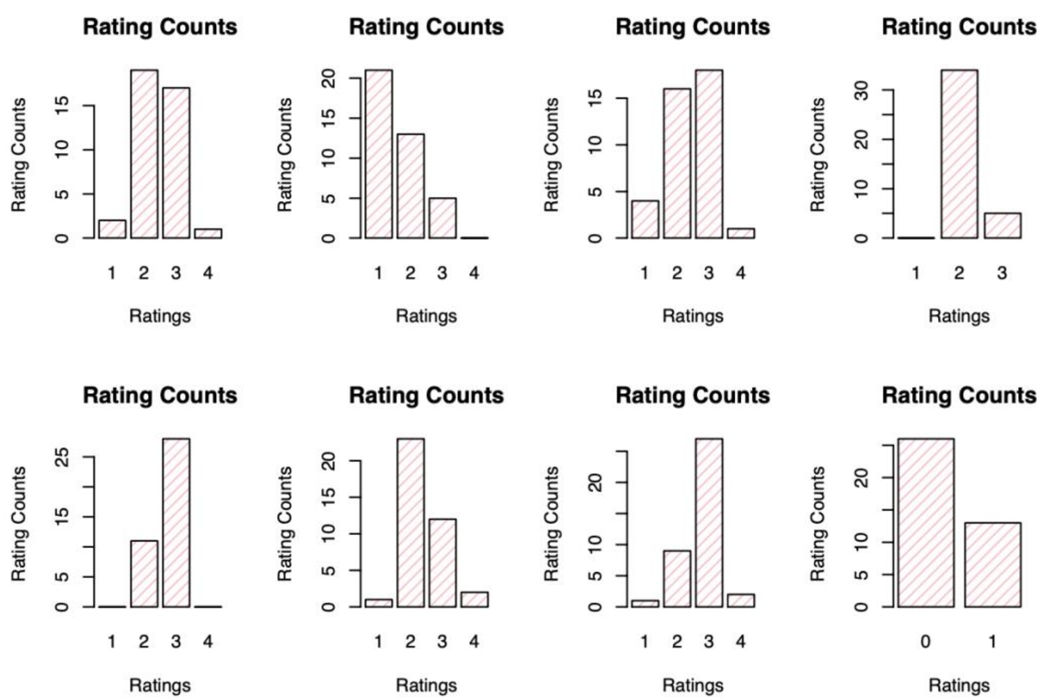
```
as.numeric(Rating) ~ (0 + Rubric | Artifact) + (0 + as.factor(Rater) |
    Artifact) + as.factor(Rater) + Semester + Rubric + as.factor(Rater):Rubric

 Groups     Name              Std.Dev. Corr
 Artifact   RubricCritDes     0.70461
            RubricInitEDA     0.56379   0.318
            RubricInterpRes   0.31946   0.142  0.674
            RubricRsrchQ      0.42310   0.500  0.194  0.538
            RubricSelMeth     0.19557   0.145  0.226  0.376 -0.240
            RubricTxtOrg      0.50026   0.268  0.437  0.364  0.305  0.213
            RubricVisOrg      0.48201   0.175  0.504  0.445  0.276 -0.161
 Artifact.1 as.factor(Rater)1 0.11320
            as.factor(Rater)2 0.33427  -0.486
            as.factor(Rater)3 0.30681   0.332  0.663
 Residual                     0.36699
```

	Estimate	Std. Error	t value
(Intercept)	1.7575545	0.11404151	15.4115336
as.factor(Rater)2	0.3660542	0.13918252	2.6300297
as.factor(Rater)3	0.1959088	0.12966636	1.5108686
SemesterS19	-0.1591805	0.07647529	-2.0814634
RubricInitEDA	0.7394940	0.12996076	5.6901329
RubricInterpRes	0.9915148	0.12770767	7.7639406
RubricRsrchQ	0.7261869	0.11793023	6.1577676
RubricSelMeth	0.4106797	0.12470498	3.2932102
RubricTxtOrg	1.0157815	0.12999540	7.8139797
RubricVisOrg	0.6542506	0.13353098	4.8996162
as.factor(Rater)2:RubricInitEDA	-0.2998076	0.15609075	-1.9207264
as.factor(Rater)3:RubricInitEDA	-0.2947319	0.15635201	-1.8850532
as.factor(Rater)2:RubricInterpRes	-0.5132297	0.15348482	-3.3438467
as.factor(Rater)3:RubricInterpRes	-0.7148433	0.15363960	-4.6527283
as.factor(Rater)2:RubricRsrchQ	-0.4874137	0.14722146	-3.3107521
as.factor(Rater)3:RubricRsrchQ	-0.3223799	0.14726517	-2.1891116
as.factor(Rater)2:RubricSelMeth	-0.3863739	0.15030941	-2.5705236
as.factor(Rater)3:RubricSelMeth	-0.3871581	0.14961457	-2.5877033
as.factor(Rater)2:RubricTxtOrg	-0.5510439	0.15646043	-3.5219379
as.factor(Rater)3:RubricTxtOrg	-0.4448937	0.15673122	-2.8385772
as.factor(Rater)2:RubricVisOrg	-0.1048994	0.15861081	-0.6613632
as.factor(Rater)3:RubricVisOrg	-0.2752130	0.15884865	-1.7325485

Our final model after adding possible random effects and model selection is as above.

Fourth research question:

The EDA bar plots for full dataset is as follows:



(Need correlation plots/ facet plots here)

Discussion:

In the first question, we notice that if we are doing barplot for the full dataset, the distribution would not make sense for some of variables but gives us a sense of how the variables are distributed. InitEDA, RsrchQ, InterpRes, VisOrg, TetOrg, SelMeth have high values of rate 3 and rate 4, and very few rate 1 and rate 2. CritDes only has roughly all numbers of rate 1. After dividing our data into subsets, we notice that there are big differences across different rubrics and raters.

For the second question, a sample interpretation for 2-way table in result section is, we see that the rater 1 and rater 2 for the rubric RsrchQ have the same rate in 5 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. Only 1 out of 13 of therates are very different. Therefore, we know that for RsrchQ, raters 1 and 2 has not much differences between each other. We see that the rater 2 and rater 3 for the rubric RsrchQ have the same rate in 7 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. 0 out of 13 of therates are very different. Therefore, we know that for RsrchQ, raters 3 and 2 has not much differences between each other.

For the third question part 3-4, the coefficients are significant if t-values in model is larger than 1.96. it does look as if the 3 raters have different ways of scoring the 7 rubrics, so the interaction we found in comb.inter_elim makes sense. Finally, we consider adding random effects to what seems like the best model so far, comb.inter_elim...The fixed-effects terms we have to work with are: as.factor(Rater),Semester and as.factor(Rater):Rubric.
In all cases, there is more than one random effect to test (3 for raters,2 for semesters, 7 for rubrics, and 21 for the interaction). We inspect AIC and BIC from anova() tables.If we accept comb.final as our final model, we can interpret the pieces as follows:

(0 + as.factor(Rater) | Artifact) + as.factor(Rater)
There is a kind of Rater x Artifact interaction: each Rater's rating on each Artifact differs from what we would expect (from the fixed effects alone) by a small random effect that depends on the Artifact

Rubric + as.factor(Rater) + as.factor(Rater):Rubric
There is a Rater x Rubric interaction: each Rater uses each Rubric in a way that is not like, or even parallel to, other rater's Rubric usage. (we saw that in the facets plot above also).

(0 + Rubric | Artifact) + Rubric

There is a kind of Rubric x Artifact interaction: There are different average scores on each rubric, but the rubric averages also vary a bit from one Artifact to the next, by a small random effect that depends on Artifact. In all of this, the fact that Rubric scores depend on Artifact (that is, there is a kind of Rubric x Artifact interaction) is what we might expect:the artifacts aren't all of equal quality on each rubric, and so we should expect the average scores on each Rubric to vary from one Artifact to the next.

More troubling are the Rater x Rubric interaction and the "kind of" Rater x Artifact interaction. The Rater x Rubric interaction suggest that the Raters are not all interpreting the Rubrics in the same way. The "kind of" Rater x Artifact interaction suggests that the Raters are not interpreting the evidence in the artifacts in the same way. These interactions suggest that perhaps the raters should be trained more, to make the raters' ratings more similar to each other.

(Need more contents here)

References:

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

Technical Appendix:

(Still needs modification, will complete in final draft)

Research Question 1

```
summary(tall)
```

```
##        X             Rater     Artifact           Repeated     
##  Min.   :  1.0   Min.   :1   Length:819         Min.   :0.0000  
##  1st Qu.:205.5   1st Qu.:1   Class :character   1st Qu.:0.0000  
##  Median :410.0   Median :2   Mode  :character   Median :0.0000  
##  Mean   :410.0   Mean   :2                      Mean   :0.3333  
##  3rd Qu.:614.5   3rd Qu.:3                      3rd Qu.:1.0000  
##  Max.   :819.0   Max.   :3                      Max.   :1.0000  
##                                                                 
##    Semester             Sex               Rubric              Rating     
##  Length:819         Length:819         Length:819         Min.   :1.000  
##  Class :character   Class :character   Class :character   1st Qu.:2.000  
##  Mode  :character   Mode  :character   Mode  :character   Median :2.000  
##                                                           Mean   :2.318  
##                                                           3rd Qu.:3.000  
##                                                           Max.   :4.000  
##                                                           NA's   :2
```

```
#Looking at the percentage from ratings:
ratings <- read.csv("ratings.csv")
summary(ratings)
```

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

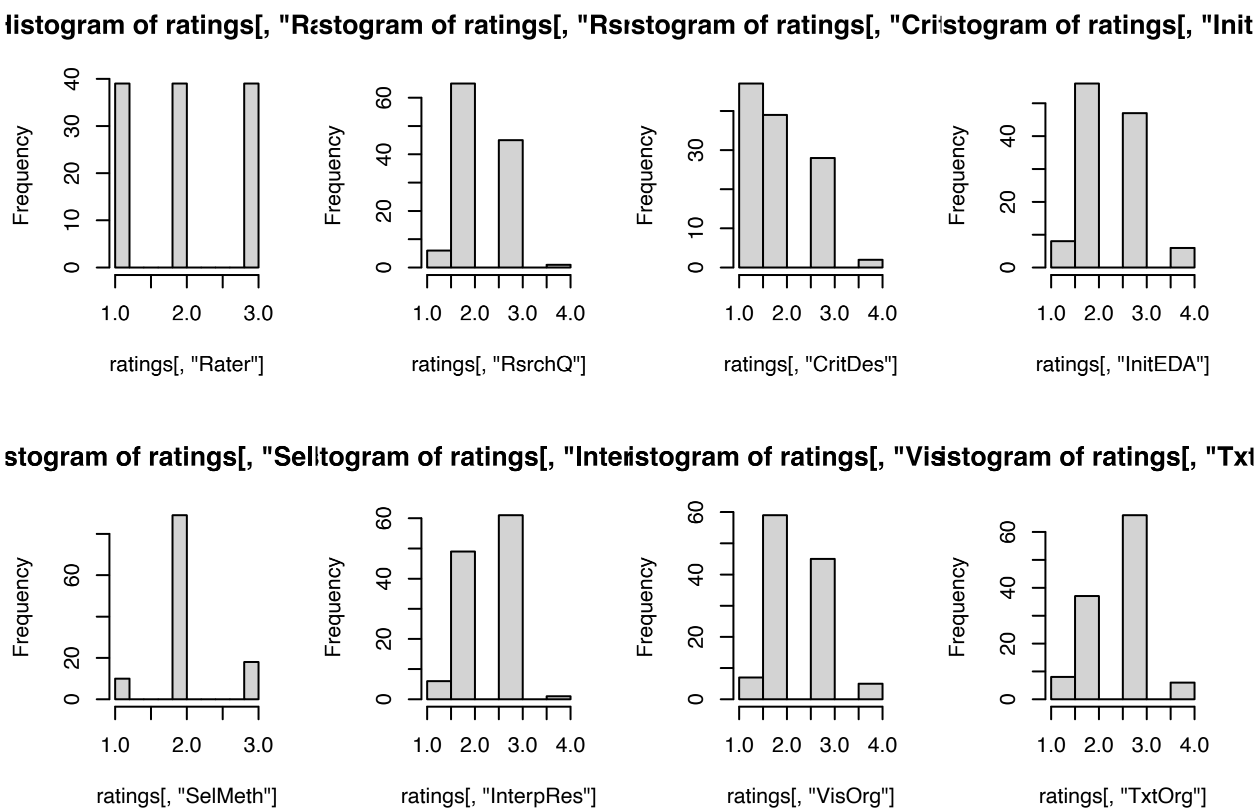
Table 1:

X	Rater	Sample	Overlap	Semester	Sex	RsrchQ	CritDes	InitEDA	SelMeth
1	3	1	5	Fall	M	3	3	2	2
2	3	2	7	Fall	F	3	3	3	3
3	3	3	9	Spring	F	2	1	3	2
4	3	4	8	Spring	M	2	2	2	1
5	3	5	NA	Fall	–	3	3	3	3
6	3	6	NA	Fall	M	2	1	2	2

```
##                      Max.   :1.0000
##
```

```
head(ratings[,1:10]) %>% kbl(booktabs=T,caption=" ") %>% kable_classic()
```

```
#Looking at the histograms of numerical variables from ratings
par(mfrow = c(2,4))
hist(ratings[,"Rater"])
hist(ratings[,"RsrchQ"])
hist(ratings[,"CritDes"])
hist(ratings[,"InitEDA"])
hist(ratings[,"SelMeth"])
hist(ratings[,"InterpRes"])
hist(ratings[,"VisOrg"])
hist(ratings[,"TxtOrg"])
```



The distribution does not make sense for some of variables but gives us a sense of how the variables are distributed.InitEDA, RsrchQ, InterpRes, VisOrg, TetOrg, SelMeth have high values of rate 3 and rate 4, and very few rate 1 and rate 2. CritDes only has roughly all numbers of rate 1.

```
#Looking at the percentage:
CritDes<-table(ratings$CritDes)
addmargins(CritDes)
```

```
##
##   1   2   3   4 Sum
##  47  39  28   2 116
```

```
round(prop.table(CritDes)*100,digits=0)
```

```
##
##  1  2  3  4
## 41 34 24  2
```

```
InitEDA <- table(ratings$InitEDA)
addmargins(InitEDA)
```

```
##
##   1   2   3   4 Sum
##   8  56  47   6 117
```

```
round(prop.table(InitEDA)*100,digits=0)
```

```
##
##  1  2  3  4
##  7 48 40  5
```

```
SelMeth <- table(ratings$SelMeth)
addmargins(SelMeth)
```

```
##
##   1   2   3 Sum
##  10  89  18 117
```

```
round(prop.table(SelMeth)*100,digits=0)
```

```
##
##  1  2  3
##  9 76 15
```

```
InterpRes <- table(ratings$InterpRes)
addmargins(InterpRes)
```

```
##
##   1   2   3   4 Sum
##   6  49  61   1 117
```

```
round(prop.table(InterpRes)*100,digits=0)
```

```
##
##  1  2  3  4
##  5 42 52  1
```

```
VisOrg <- table(ratings$VisOrg)
addmargins(VisOrg)
```

```
##
##   1   2   3   4 Sum
```

```
##   7  59  45   5 116
```

```
round(prop.table(VisOrg)*100,digits=0)
```

```
##
##  1  2  3  4
##  6 51 39  4
```

```
TxtOrg <- table(ratings$TxtOrg)
addmargins(TxtOrg)
```

```
##
##   1   2   3   4 Sum
##   8  37  66   6 117
```

```
round(prop.table(TxtOrg)*100,digits=0)
```

```
##
##  1  2  3  4
##  7 32 56  5
```

```
Artifact <- table(ratings$Artifact)
addmargins(Artifact)
```

```
##
## 100 101 102 103 104 105 106 107 111 112 113 114 115 116 117 118  13  15  16  17
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  21  22  23  24  25  26  27  28  32  33  34  35  36  37  38  39  40  45  46  47
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  48  49   5  53  54  55  56  57   6  61  62  63  64  65  66  67  68   7  72  73
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  74  75  76  77  78  79   8  84  85  86  87  88   9  92  93  94  95  96  01 010
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   3   3
## 011 012 013  02  03  04  05  06  07  08  09 Sum
##   3   3   3   3   3   3   3   3   3   3   3 117
```

```
round(prop.table(Artifact)*100,digits=0)
```

```
##
## 100 101 102 103 104 105 106 107 111 112 113 114 115 116 117 118  13  15  16  17
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  21  22  23  24  25  26  27  28  32  33  34  35  36  37  38  39  40  45  46  47
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  48  49   5  53  54  55  56  57   6  61  62  63  64  65  66  67  68   7  72  73
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  74  75  76  77  78  79   8  84  85  86  87  88   9  92  93  94  95  96  01 010
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   3   3
## 011 012 013  02  03  04  05  06  07  08  09
##   3   3   3   3   3   3   3   3   3   3   3
```

```
Repeated <- table(ratings$Repeated)
addmargins(Repeated)
```

```
##
##   0   1 Sum
##  78  39 117
```

```
round(prop.table(Repeated)*100,digits=0)
```

```
##
##  0  1
## 67 33
```

```
RsrchQ <- table(ratings$RsrchQ)
addmargins(RsrchQ)
```

```
##
##   1   2   3   4 Sum
##   6  65  45   1 117
```

```
round(prop.table(RsrchQ)*100,digits=0)
```

```
##
##  1  2  3  4
##  5 56 38  1
```

```
Sex <- table(ratings$Sex)
addmargins(Sex)
```

```
##
##  --   F   M Sum
##   1  64  52 117
```

```
round(prop.table(Sex)*100,digits=0)
```

```
##
## --  F  M
##  1 55 44
```

```
Semester <- table(ratings$Semester)
addmargins(Semester)
```

```
##
##   Fall Spring    Sum
##     83     34    117
```

```
round(prop.table(Semester)*100,digits=0)
```

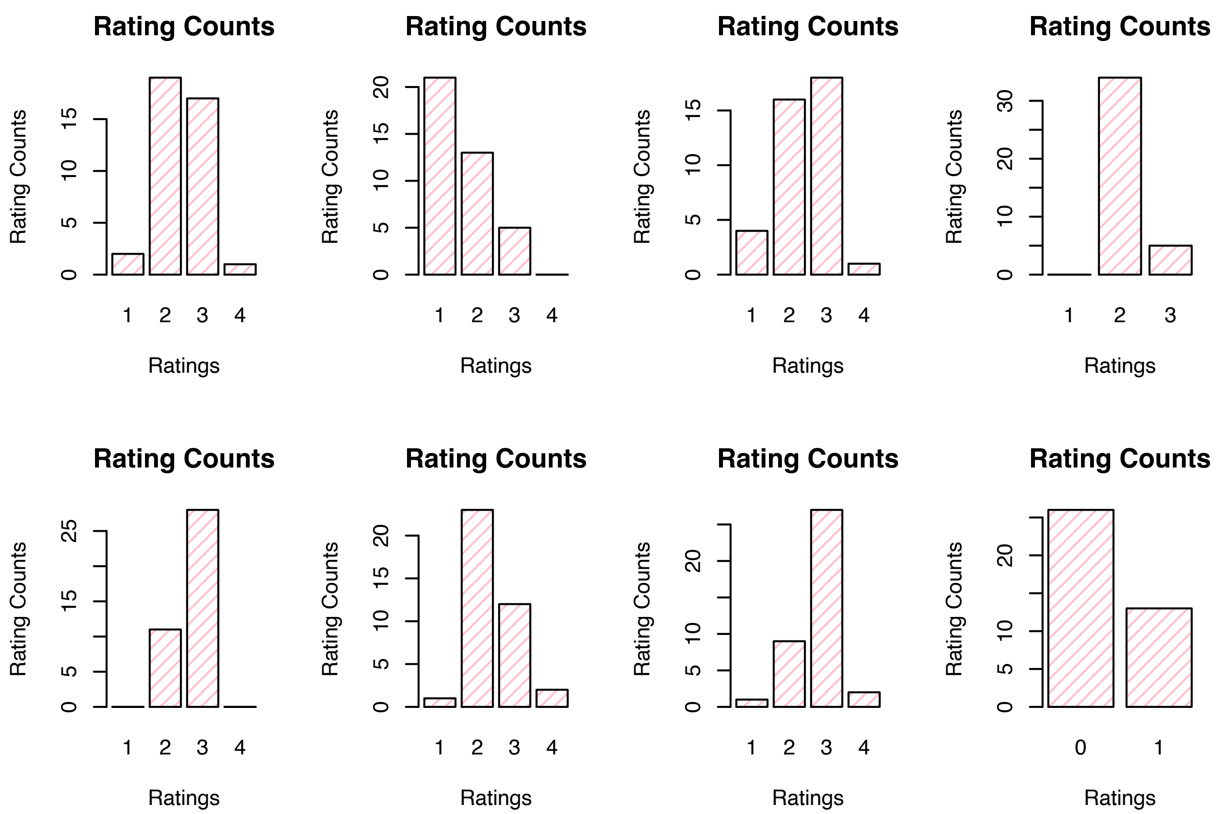
```
##
##   Fall Spring
##     71     29
```

These percentage tables help us look at the exact percentages of each factors.

```
ratings$X <- as.factor(ratings$X)
ratings$Rater <- as.factor(ratings$Rater)
ratings$Sample <- as.factor(ratings$Sample)
ratings$Overlap <- as.factor(ratings$Overlap)
ratings$Semester <- as.factor(ratings$Semester)
ratings$Sex<- as.factor(ratings$Sex)
ratings$RsrchQ <- as.factor(ratings$RsrchQ)
ratings$CritDes <- as.factor(ratings$CritDes)
ratings$InitEDA <- as.factor(ratings$InitEDA)
ratings$SelMeth <- as.factor(ratings$SelMeth)
ratings$InterpRes <- as.factor(ratings$InterpRes)
ratings$VisOrg <- as.factor(ratings$VisOrg)
ratings$TxtOrg <- as.factor(ratings$TxtOrg)
ratings$Artifact <- as.factor(ratings$Artifact)
```

```r
ratings$Repeated <- as.factor(ratings$Repeated)
```

```r
#Looking at the barplot of rating score=1 as an example:
par(mfrow=c(2,4))
ratings_1<- ratings %>% filter(ratings$Rater==1)
barplot(table(ratings_1$RsrchQ),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$CritDes),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$InitEDA),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$SelMeth),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$InterpRes),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink

barplot(table(ratings_1$VisOrg),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$TxtOrg),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$Repeated),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink"
```



Looking at the bar plots, we see that there are differences in rating counts when rating score=1. If we do a further EDA on other rating scores, the result would probably be the same due to the fact that the rating counts for different groups are very different.

Research Question 2

```r
# Measuring the correlations to see if raters agree with each other
common <- tall[grep("0",tall$Artifact),]
```

```r
head(common)
```

```
##     X Rater Artifact Repeated Semester Sex Rubric Rating
## 1   1     3       O5        1      F19   M RsrchQ      3
## 2   2     3       O7        1      F19   F RsrchQ      3
## 3   3     3       O9        1      S19   F RsrchQ      2
## 4   4     3       O8        1      S19   M RsrchQ      2
## 10 10     3      O10        1      F19   F RsrchQ      2
## 11 11     3      O13        1      F19   M RsrchQ      2
```

```r
CritDes.ratings <- common[common$Rubric=="CritDes",]
InitEDA.ratings <- common[common$Rubric=="InitEDA",]
SelMeth.ratings <- common[common$Rubric=="SelMeth",]
InterpRes.ratings <- common[common$Rubric=="InterpRes",]
VisOrg.ratings <- common[common$Rubric=="VisOrg",]
TxtOrg.ratings <- common[common$Rubric=="TxtOrg",]
```

```r
CritDes_m=lmer(Rating ~ 1 + (1|Artifact), data=CritDes.ratings)
InitEDA_m=lmer(Rating ~ 1 + (1|Artifact), data=InitEDA.ratings)
SelMeth_m=lmer(Rating ~ 1 + (1|Artifact), data=SelMeth.ratings)
InterpRes_m=lmer(Rating ~ 1 + (1|Artifact), data=InterpRes.ratings)
VisOrg_m=lmer(Rating ~ 1 + (1|Artifact), data=VisOrg.ratings)
TxtOrg_m=lmer(Rating ~ 1 + (1|Artifact), data=TxtOrg.ratings)
```

```r
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_RsrchQ <- data.frame(
  r1=repeated$RsrchQ[repeated$Rater==1],r2=repeated$RsrchQ[repeated$Rater==2],a1=repeated$Artifact[re
r1 <- factor(raters_1_and_2_on_RsrchQ$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_RsrchQ$r2,levels=1:4)
(t12 <- 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
```

We see that the rater 1 and rater 2 for the rubric RsrchQ have the same rate in 5 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. Only 1 out of 13 of therates are very different. Therefore, we know that for RsrchQ, raters 1 and 2 has not much differences between each other.

```r
raters_2_and_3_on_RsrchQ <- data.frame(r2=repeated$RsrchQ[repeated$Rater==2],r3=repeated$RsrchQ[repea
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3]
)
r2 <- factor(raters_2_and_3_on_RsrchQ$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_RsrchQ$r3,levels=1:4)
(t23 <- 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
```

We see that the rater 2 and rater 3 for the rubric RsrchQ have the same rate in 7 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. 0 out of 13 of therates are very different. Therefore, we know that for RsrchQ, raters 3 and 2 has not much differences between each other.

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

```
## 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
```

```
#Calculating ICC value
performance::icc(model=RsrchQ_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.189
##   Conditional ICC: 0.189
```

```
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_CritDes <- data.frame(r1=repeated$CritDes[repeated$Rater==1],r2=repeated$CritDes[re
a2=repeated$Artifact[repeated$Rater==2]
)
r1 <- factor(raters_1_and_2_on_CritDes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_CritDes$r2,levels=1:4)
(t12 <- 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
```

We see that the rater 1 and rater 2 for the rubric CritDes have the same rate in 7 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. Only 1 out of 13 of therates are very different. Therefore, we know that for CritDes, raters 1 and 2 has not much differences between each other.

```r
raters_2_and_3_on_CritDes <- data.frame(r2=repeated$CritDes[repeated$Rater==2],r3=repeated$CritDes[re
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3])
r2 <- factor(raters_2_and_3_on_CritDes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_CritDes$r3,levels=1:4)
(t23 <- 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
```

We see that the rater 3 and rater 2 for the rubric CritDes have the same rate in 9 out of 13 of the cases. Even for some artifacts they had different rates, most of them are different by 2 and 3. Only 1 out of 13 of therates are very different. Therefore, we know that for CritDes, raters 3 and 2 has not much differences between each other.

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

```
## 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
```

```r
#Icc value:
performance::icc(model=CritDes_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.573
##   Conditional ICC: 0.573
```

Lower ICC means less agreement accross groups, so here we have higher agreements.

```r
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_InitEDA <- data.frame(r1=repeated$InitEDA[repeated$Rater==1],
r2=repeated$InitEDA[repeated$Rater==2], a1=repeated$Artifact[repeated$Rater==1],
a2=repeated$Artifact[repeated$Rater==2]
)
```

```
r1 <- factor(raters_1_and_2_on_InitEDA$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InitEDA$r2,levels=1:4)
(t12 <- 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
```

For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric InitEDA have the same rate in 9/13 of the cases, even for some artifacts they had different rates, most of them are only r1 = 2 and r2 = 3 or r2 = 2 and r1 = 3 (i.e. the rates are not that different).

```
raters_2_and_3_on_InitEDA <- data.frame(r2=repeated$InitEDA[repeated$Rater==2], r3=repeated$InitEDA[r
)
r2 <- factor(raters_2_and_3_on_InitEDA$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InitEDA$r3,levels=1:4)
(t23 <- 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
```

We find that the rater 2 and rater 3 for the rubric InitEDA have the same rate in 9/13 of the cases, even for some artifacts they had different rates, most of them are only r1 = 2 and r2 = 3 or r2 = 2 and r1 = 3 (i.e. the rates are not that different).So for the rubric InitEDA, the rater 2 and 3, they do not usually disagree with each other.

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

```
## 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
```

```
performance::icc(InitEDA_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.493
##   Conditional ICC: 0.493
```

```
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_SelMeth <- data.frame(r1=repeated$SelMeth[repeated$Rater==1],
r2=repeated$SelMeth[repeated$Rater==2], a1=repeated$Artifact[repeated$Rater==1],
a2=repeated$Artifact[repeated$Rater==2]
)
r1 <- factor(raters_1_and_2_on_SelMeth$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_SelMeth$r2,levels=1:4)
(t12 <- 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
```

We find that the rater 1 and rater 2 for the rubric SelMeth have the same rate in 12/13 of the cases, even for the rest of the artifact it had different rates, it is only r2 = 1 and r1 = 2.So for the rubric SelMeth, the rater 1 and 2, they do not usually disagree with each other.

```
raters_2_and_3_on_SelMeth <- data.frame(r2=repeated$SelMeth[repeated$Rater==2], r3=repeated$SelMeth[r
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3]
)
r2 <- factor(raters_2_and_3_on_SelMeth$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_SelMeth$r3,levels=1:4)
(t23 <- 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
```

We find that the rater 2 and rater 3 for the rubric SelMeth have the same rate in 9/13 of the cases, even for some artifacts they had different rates, most of them are only r1 = 2 and r2 = 3 or r2 = 2 and r1 = 3 So for the rubric SelMeth, the rater 2 and 3, they do not usually disagree with each other.

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

```
## 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
```

```
performance::icc(SelMeth_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.521
##   Conditional ICC: 0.521
```

```
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_InterpRes <- data.frame(r1=repeated$InterpRes[repeated$Rater==1],
r2=repeated$InterpRes[repeated$Rater==2], a1=repeated$Artifact[repeated$Rater==1],
a2=repeated$Artifact[repeated$Rater==2])
r1 <- factor(raters_1_and_2_on_InterpRes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InterpRes$r2,levels=1:4)
(t12 <- 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
```

We find that the rater 1 and rater 2 for the rubric InterpRes have the same rate in 8/13 of the cases, even for the rest of the artifact it had different rates, it is only r2 = 1 and r1 = 2. Only one of the artifact had |r1-r2|=2.

```
raters_2_and_3_on_InterpRes <- data.frame(r2=repeated$InterpRes[repeated$Rater==2], r3=repeated$Inter
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3]
)
r2 <- factor(raters_2_and_3_on_InterpRes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InterpRes$r3,levels=1:4)
(t23 <- 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
```

We find that the rater 2 and rater 3 for the rubric InterpRes have the same rate in 8/13 of the cases, even for some artifacts they had different rates, most of them are 2 and 3. Only one of the artifact had |r1-r2|=2.So for the rubric InterpRes, the rater 2 and 3, they do not usually disagree with each other.

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

```
## 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
```

```
performance::icc(InterpRes_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.230
##   Conditional ICC: 0.230
```

```
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_VisOrg <- data.frame(r1=repeated$VisOrg[repeated$Rater==1],
r2=repeated$VisOrg[repeated$Rater==2], a1=repeated$Artifact[repeated$Rater==1],
a2=repeated$Artifact[repeated$Rater==2]
)
r1 <- factor(raters_1_and_2_on_VisOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_VisOrg$r2,levels=1:4)
(t12 <- 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
```

```
raters_2_and_3_on_VisOrg <- data.frame(r2=repeated$VisOrg[repeated$Rater==2], r3=repeated$VisOrg[repe
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3]
)
r2 <- factor(raters_2_and_3_on_VisOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_VisOrg$r3,levels=1:4)
(t23 <- 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
```

We find that the rater 1 and rater 2 for the rubric VisOrg have the same rate in 7/13 of the cases, even for the rest of the artifact it had different rates, it is only r2 = 1 and r1 = 2. Only one of the artifact had |r1-r2|=2.We find that the rater 2 and rater 3 for the rubric VisOrg have the same rate in 8/13 of the cases, even for some artifacts they had different rates, most of them are 2 adn 3. So for the rubric VisOrg, the rater 1 and 2, they do not usually disagree with each other.

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

```
## 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
```

```
performance::icc(VisOrg_m)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.592
##   Conditional ICC: 0.592
```

```
repeated <- ratings[ratings$Repeated==1,]
raters_1_and_2_on_TxtOrg <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
r2=repeated$TxtOrg[repeated$Rater==2], a1=repeated$Artifact[repeated$Rater==1],
a2=repeated$Artifact[repeated$Rater==2]
)
r1 <- factor(raters_1_and_2_on_TxtOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_TxtOrg$r2,levels=1:4)
(t12 <- 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
```

We find that the rater 1 and rater 2 for the rubric TxtOrg have the same rate in 9/13 of the cases, even for the rest of the artifact it had different rates, it is only r2 = 1 and r1 = 2. Only one of the artifact had |r1-r2|=2.

```r
raters_2_and_3_on_TxtOrg <- data.frame(r2=repeated$TxtOrg[repeated$Rater==2], r3=repeated$TxtOrg[repe
a1=repeated$Artifact[repeated$Rater==2], a2=repeated$Artifact[repeated$Rater==3]
)
r2 <- factor(raters_2_and_3_on_TxtOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_TxtOrg$r3,levels=1:4)
(t23 <- 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
```

We find that the rater 2 and rater 3 for the rubric TxtOrg have the same rate in 7/13 of the cases, even for some artifacts they had different rates, most of them are 2 and 3. Only onehad |r1-r2|=2.So for the rubric TxtOrg, the rater 2 and 3, they do not usually disagree with each other.

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

```
## 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
```

```r
performance::icc(TxtOrg_m)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.143
##   Conditional ICC: 0.143
```

We found that fitting the lmer model to (1|Raters) does not work here due to the singularity reason. It was our first approach but since we can't get a good ICC value, I prefer models grouped by artifacts.

Research Question 3

```
tall$Rater = as.factor(tall$Rater)
m1 = lmer(Rating ~ 1 + Rater + (0+Rubric | Artifact),data = tall)
summary(m1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1478.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1649 -0.5068 -0.0540  0.5366  3.6782
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.6708   0.8190
##           RubricInitEDA   0.3714   0.6095   0.28
##           RubricInterpRes 0.2166   0.4654   0.00 0.78
##           RubricRsrchQ    0.1625   0.4031   0.42 0.48 0.71
##           RubricSelMeth   0.1047   0.3235   0.61 0.36 0.36 0.25
##           RubricTxtOrg    0.3711   0.6092   0.03 0.67 0.78 0.61 0.20
##           RubricVisOrg    0.2911   0.5396   0.19 0.77 0.74 0.56 0.26 0.77
##  Residual                 0.1925   0.4387
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  2.316335   0.050594  45.783
## Rater2       0.002191   0.055501   0.039
## Rater3      -0.164062   0.055500  -2.956
##
## Correlation of Fixed Effects:
##        (Intr) Rater2
## Rater2 -0.549
## Rater3 -0.549  0.500
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00429684 (tol = 0.002, component 1)
```

```
m2 = lmer(Rating ~ 1 + Rubric + (0+Rubric | Artifact),data = tall)
```

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

```
summary(m2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1439.7
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9610 -0.5301 -0.0116  0.5062  3.8765
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes    0.54648  0.7392
##           RubricInitEDA    0.34214  0.5849   0.47
##           RubricInterpRes  0.19783  0.4448   0.26 0.76
##           RubricRsrchQ     0.16164  0.4020   0.59 0.45 0.72
##           RubricSelMeth    0.08874  0.2979   0.43 0.63 0.79 0.48
##           RubricTxtOrg     0.27144  0.5210   0.35 0.64 0.74 0.59 0.71
##           RubricVisOrg     0.28462  0.5335   0.37 0.75 0.73 0.56 0.53 0.79
##  Residual                  0.19251  0.4388
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      1.90899    0.08907  21.432
## RubricInitEDA    0.54139    0.09491   5.704
## RubricInterpRes  0.58064    0.09999   5.807
## RubricRsrchQ     0.45341    0.08680   5.224
## RubricSelMeth    0.15830    0.09228   1.715
## RubricTxtOrg     0.68645    0.09870   6.955
## RubricVisOrg     0.52298    0.09813   5.329
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO
## RubrcIntEDA -0.673
## RbrcIntrpRs -0.787  0.734
## RubrcRsrchQ -0.771  0.589  0.751
## RubricSlMth -0.836  0.667  0.775  0.690
## RubrcTxtOrg -0.734  0.677  0.752  0.683  0.729
## RubricVsOrg -0.722  0.718  0.744  0.667  0.681  0.752
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
m3 = lmer(Rating ~ 1 + Sex + (0+Rubric | Artifact),data = tall)
summary(m3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Sex + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1480.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0275 -0.4958 -0.0692  0.5067  3.7970
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes    0.63000  0.7937
##           RubricInitEDA    0.38328  0.6191   0.26
##           RubricInterpRes  0.26165  0.5115   0.00 0.79
```

```
##            RubricRsrchQ    0.17971  0.4239   0.38 0.51 0.75
##            RubricSelMeth   0.09281  0.3046   0.55 0.38 0.42 0.27
##            RubricTxtOrg    0.40787  0.6386   0.03 0.69 0.81 0.65 0.25
##            RubricVisOrg    0.32590  0.5709   0.18 0.79 0.78 0.61 0.31 0.80
##  Residual                  0.19324  0.4396
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)    3.0000     0.3933   7.628
## SexF          -0.8000     0.3970  -2.015
## SexM          -0.7403     0.3979  -1.860
##
## Correlation of Fixed Effects:
##      (Intr) SexF
## SexF -0.991
## SexM -0.988  0.979
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00739175 (tol = 0.002, component 1)
```

```
m4 = lmer(Rating ~ 1 + Rubric + Rater + (0+Rubric | Artifact),data = tall)
```

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

```
summary(m4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1436.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0997 -0.5121  0.0013  0.5138  3.7700
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.55942  0.7479
##           RubricInitEDA   0.34207  0.5849   0.47
##           RubricInterpRes 0.16894  0.4110   0.25 0.75
##           RubricRsrchQ    0.15414  0.3926   0.60 0.43 0.69
##           RubricSelMeth   0.08392  0.2897   0.45 0.62 0.76 0.44
##           RubricTxtOrg    0.25881  0.5087   0.35 0.62 0.71 0.56 0.69
##           RubricVisOrg    0.26581  0.5156   0.37 0.74 0.70 0.53 0.49 0.77
##  Residual                 0.19069  0.4367
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)     1.965e+00  9.528e-02  20.621
## RubricInitEDA   5.413e-01  9.489e-02   5.704
## RubricInterpRes 5.802e-01  1.000e-01   5.801
## RubricRsrchQ    4.533e-01  8.680e-02   5.222
## RubricSelMeth   1.582e-01  9.230e-02   1.714
```

```
## RubricTxtOrg     6.860e-01  9.873e-02   6.948
## RubricVisOrg     5.230e-01  9.817e-02   5.327
## Rater2          -9.092e-05  5.535e-02  -0.002
## Rater3          -1.668e-01  5.535e-02  -3.014
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO Rater2
## RubrcIntEDA -0.636
## RbrcIntrpRs -0.760  0.734
## RubrcRsrchQ -0.735  0.588  0.754
## RubricSlMth -0.793  0.667  0.780  0.690
## RubrcTxtOrg -0.701  0.676  0.752  0.683  0.730
## RubricVsOrg -0.693  0.717  0.745  0.667  0.683  0.751
## Rater2      -0.290  0.000  0.000  0.000  0.000  0.000 -0.001
## Rater3      -0.290  0.000  0.000  0.000  0.000  0.000 -0.001  0.500
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```r
m5 = lmer(Rating ~ 1 + Rater + Sex + (0+Rubric | Artifact),data = tall)
```

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

```r
summary(m5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + Sex + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1476.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1755 -0.4955 -0.0530  0.5412  3.6881
##
## Random effects:
##  Groups   Name              Variance Std.Dev. Corr
##  Artifact RubricCritDes     0.64941  0.8059
##           RubricInitEDA     0.37420  0.6117    0.26
##           RubricInterpRes   0.22184  0.4710   -0.03  0.78
##           RubricRsrchQ      0.16590  0.4073    0.40  0.49  0.72
##           RubricSelMeth     0.09161  0.3027    0.58  0.34  0.34  0.22
##           RubricTxtOrg      0.38141  0.6176    0.02  0.67  0.79  0.62  0.19
##           RubricVisOrg      0.29791  0.5458    0.17  0.77  0.75  0.57  0.24  0.78
##  Residual                   0.19151  0.4376
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  3.172324   0.381490   8.316
## Rater2       0.003592   0.055176   0.065
## Rater3      -0.172324   0.055335  -3.114
## SexF        -0.900284   0.382447  -2.354
## SexM        -0.843507   0.383143  -2.202
##
## Correlation of Fixed Effects:
```

```
##        (Intr) Rater2 Rater3 SexF
## Rater2 -0.072
## Rater3 -0.145  0.499
## SexF   -0.988  0.003  0.077
## SexM   -0.985 -0.003  0.069  0.979
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
m6 = lmer(Rating ~ 1 + Rubric + Sex + (0+Rubric | Artifact),data = tall)
summary(m6)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Sex + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1438.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9634 -0.5238 -0.0149  0.4950  3.8926
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.53569  0.7319
##           RubricInitEDA   0.34252  0.5853   0.46
##           RubricInterpRes 0.20086  0.4482   0.25 0.75
##           RubricRsrchQ    0.16499  0.4062   0.58 0.44 0.73
##           RubricSelMeth   0.08221  0.2867   0.40 0.62 0.79 0.47
##           RubricTxtOrg    0.27633  0.5257   0.34 0.63 0.74 0.59 0.71
##           RubricVisOrg    0.29069  0.5392   0.37 0.75 0.73 0.57 0.53 0.79
##  Residual                 0.19147  0.4376
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      2.65225    0.40255   6.589
## RubricInitEDA    0.54139    0.09491   5.704
## RubricInterpRes  0.58077    0.09999   5.808
## RubricRsrchQ     0.45383    0.08678   5.230
## RubricSelMeth    0.15869    0.09245   1.717
## RubricTxtOrg     0.68637    0.09872   6.953
## RubricVisOrg     0.52329    0.09815   5.331
## SexF            -0.77794    0.39877  -1.951
## SexM            -0.71864    0.39967  -1.798
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO SexF
## RubrcIntEDA -0.147
## RbrcIntrpRs -0.172  0.734
## RubrcRsrchQ -0.166  0.588  0.752
## RubricSlMth -0.180  0.666  0.774  0.690
## RubrcTxtOrg -0.160  0.676  0.752  0.683  0.728
## RubricVsOrg -0.157  0.718  0.745  0.667  0.680  0.752
## SexF        -0.972  0.000  0.000 -0.002 -0.004  0.000 -0.001
## SexM        -0.969  0.000  0.000 -0.002 -0.004  0.000 -0.001  0.979
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00867034 (tol = 0.002, component 1)
```

```
m7 = lmer(Rating ~ 1 + Rubric + Rater + Repeated + (0+Rubric | Artifact),data = tall)
summary(m7)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Repeated + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1438.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0933 -0.5159 -0.0002  0.5202  3.7837
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.55709  0.7464
##           RubricInitEDA   0.34270  0.5854   0.47
##           RubricInterpRes 0.17298  0.4159   0.25 0.75
##           RubricRsrchQ    0.15437  0.3929   0.60 0.43 0.70
##           RubricSelMeth   0.08639  0.2939   0.44 0.62 0.77 0.44
##           RubricTxtOrg    0.26415  0.5140   0.35 0.63 0.72 0.57 0.70
##           RubricVisOrg    0.26306  0.5129   0.37 0.74 0.70 0.52 0.48 0.77
##  Residual                 0.19050  0.4365
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)      1.973222   0.096243  20.502
## RubricInitEDA    0.541271   0.094893   5.704
## RubricInterpRes  0.580829   0.100034   5.806
## RubricRsrchQ     0.455911   0.086947   5.244
## RubricSelMeth    0.162641   0.092702   1.754
## RubricTxtOrg     0.685977   0.098715   6.949
## RubricVisOrg     0.523979   0.098205   5.336
## Rater2           0.001555   0.055396   0.028
## Rater3          -0.165257   0.055395  -2.983
## Repeated        -0.066760   0.102409  -0.652
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO Rater2 Rater3
## RubrcIntEDA -0.628
## RbrcIntrpRs -0.747  0.734
## RubrcRsrchQ -0.717  0.587  0.753
## RubricSlMth -0.767  0.665  0.778  0.691
## RubrcTxtOrg -0.689  0.676  0.752  0.683  0.728
## RubricVsOrg -0.683  0.717  0.745  0.668  0.682  0.752
## Rater2      -0.287 -0.001  0.000 -0.001  0.000  0.000 -0.001
## Rater3      -0.288  0.000  0.000  0.000  0.000  0.000 -0.001  0.500
## Repeated    -0.151  0.001 -0.012 -0.051 -0.080  0.002 -0.020  0.000  0.000
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00274453 (tol = 0.002, component 1)
```

```r
m8 = lmer(Rating ~ 1 + Rater + Sex + Repeated +(0+Rubric | Artifact),data = tall)
summary(m8)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + Sex + Repeated + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1478.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1546 -0.4854 -0.0644  0.5324  3.7167
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes    0.64270  0.8017
##           RubricInitEDA    0.37639  0.6135    0.26
##           RubricInterpRes  0.23051  0.4801   -0.03  0.78
##           RubricRsrchQ     0.16806  0.4100    0.38  0.49  0.72
##           RubricSelMeth    0.08927  0.2988    0.56  0.36  0.37  0.22
##           RubricTxtOrg     0.39190  0.6260    0.03  0.68  0.80  0.63  0.23
##           RubricVisOrg     0.29528  0.5434    0.16  0.77  0.75  0.57  0.24  0.78
##  Residual                  0.19146  0.4376
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  3.169457   0.382705   8.282
## Rater2       0.006588   0.055221   0.119
## Rater3      -0.169457   0.055380  -3.060
## SexF        -0.889190   0.384125  -2.315
## SexM        -0.831833   0.384882  -2.161
## Repeated    -0.096326   0.097836  -0.985
##
## Correlation of Fixed Effects:
##          (Intr) Rater2 Rater3 SexF   SexM
## Rater2   -0.072
## Rater3   -0.145  0.499
## SexF     -0.986  0.003  0.076
## SexM     -0.983 -0.003  0.069  0.979
## Repeated  0.000  0.001 -0.003 -0.048 -0.052
```

```r
m9 = m9 = lmer(Rating ~ 1 + Rubric + Sex + Repeated +(0+Rubric | Artifact),data = tall)
```

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

```r
summary(m9)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Sex + Repeated + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1441.2
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -2.9557 -0.5235 -0.0223  0.4938  3.9028
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.53418  0.7309
##           RubricInitEDA   0.34330  0.5859   0.46
##           RubricInterpRes 0.20468  0.4524   0.25 0.75
##           RubricRsrchQ    0.16540  0.4067   0.57 0.45 0.73
##           RubricSelMeth   0.08498  0.2915   0.39 0.62 0.79 0.47
##           RubricTxtOrg    0.28140  0.5305   0.34 0.64 0.74 0.60 0.72
##           RubricVisOrg    0.28813  0.5368   0.36 0.75 0.73 0.57 0.52 0.79
##  Residual                 0.19133  0.4374
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      2.65247    0.40432   6.560
## RubricInitEDA    0.54142    0.09491   5.704
## RubricInterpRes  0.58076    0.10002   5.807
## RubricRsrchQ     0.45607    0.08690   5.248
## RubricSelMeth    0.16295    0.09285   1.755
## RubricTxtOrg     0.68615    0.09870   6.952
## RubricVisOrg     0.52380    0.09817   5.335
## SexF            -0.76876    0.40105  -1.917
## SexM            -0.70965    0.40202  -1.765
## Repeated        -0.06382    0.10518  -0.607
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO SexF   SexM
## RubrcIntEDA -0.146
## RbrcIntrpRs -0.171  0.734
## RubrcRsrchQ -0.165  0.588  0.751
## RubricSlMth -0.179  0.664  0.772  0.691
## RubrcTxtOrg -0.159  0.677  0.752  0.682  0.725
## RubricVsOrg -0.156  0.718  0.744  0.667  0.678  0.752
## SexF        -0.971  0.000  0.000  0.000  0.000  0.000  0.000
## SexM        -0.969  0.000  0.000  0.000  0.000  0.000  0.000  0.979
## Repeated     0.010  0.000 -0.001 -0.046 -0.081  0.006 -0.012 -0.047 -0.050
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```r
m10 = lmer(Rating ~ 1 + Rater + Semester+ Rubric+(0+Rubric | Artifact),data = tall)
```

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

```r
summary(m10)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + Semester + Rubric + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1435.3
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -3.1190 -0.5101 -0.0149  0.5187  3.7768
##
## Random effects:
##  Groups   Name              Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.56198  0.7497
##           RubricInitEDA   0.35041  0.5920   0.48
##           RubricInterpRes 0.17215  0.4149   0.25 0.75
##           RubricRsrchQ    0.17157  0.4142   0.60 0.45 0.72
##           RubricSelMeth   0.06985  0.2643   0.44 0.62 0.75 0.44
##           RubricTxtOrg    0.25478  0.5048   0.35 0.62 0.70 0.56 0.67
##           RubricVisOrg    0.26011  0.5100   0.37 0.74 0.69 0.53 0.44 0.76
##  Residual                 0.18909  0.4348
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)      2.022484   0.098866  20.457
## Rater2          -0.001374   0.054994  -0.025
## Rater3          -0.168826   0.054993  -3.070
## SemesterS19     -0.184472   0.084146  -2.192
## RubricInitEDA    0.541716   0.094886   5.709
## RubricInterpRes  0.580228   0.100001   5.802
## RubricRsrchQ     0.453257   0.086690   5.228
## RubricSelMeth    0.156550   0.092721   1.688
## RubricTxtOrg     0.685876   0.098769   6.944
## RubricVisOrg     0.522887   0.098209   5.324
##
## Correlation of Fixed Effects:
##             (Intr) Rater2 Rater3 SmsS19 RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO
## Rater2      -0.282
## Rater3      -0.282  0.500
## SemesterS19 -0.267  0.016  0.016
## RubrcIntEDA -0.610 -0.001  0.000 -0.001
## RbrcIntrpRs -0.732 -0.001  0.000  0.000  0.734
## RubrcRsrchQ -0.698 -0.001  0.000  0.002  0.588  0.756
## RubricSlMth -0.777  0.000  0.000  0.006  0.663  0.779  0.689
## RubrcTxtOrg -0.679 -0.001  0.000 -0.001  0.676  0.751  0.684  0.728
## RubricVsOrg -0.672 -0.001 -0.001  0.000  0.716  0.745  0.668  0.682  0.751
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```r
m11 = lmer(Rating ~ 1 + Rater + Sex + Semester+(0+Rubric | Artifact),data = tall)
```

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

```r
summary(m11)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + Sex + Semester + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1476.1
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3.1991 -0.4964 -0.0598  0.5344  3.6773
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.6505   0.8065
##           RubricInitEDA   0.3849   0.6204    0.27
##           RubricInterpRes 0.2269   0.4763   -0.02  0.78
##           RubricRsrchQ    0.1804   0.4247    0.40  0.51  0.73
##           RubricSelMeth   0.0806   0.2839    0.58  0.33  0.31  0.21
##           RubricTxtOrg    0.3810   0.6172    0.02  0.67  0.78  0.62  0.15
##           RubricVisOrg    0.2911   0.5395    0.16  0.77  0.73  0.58  0.19  0.78
##  Residual                 0.1907   0.4367
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  3.173760   0.377750   8.402
## Rater2       0.002673   0.054939   0.049
## Rater3      -0.173760   0.055098  -3.154
## SexF        -0.835671   0.380510  -2.196
## SexM        -0.823366   0.379567  -2.169
## SemesterS19 -0.169264   0.086856  -1.949
##
## Correlation of Fixed Effects:
##             (Intr) Rater2 Rater3 SexF   SexM
## Rater2      -0.073
## Rater3      -0.146  0.499
## SexF        -0.983  0.002  0.077
## SexM        -0.984 -0.004  0.069  0.978
## SemesterS19  0.000  0.008  0.000 -0.099 -0.035
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
m12 = lmer(Rating ~ 1 + Rubric + Sex + Semester+(0+Rubric | Artifact),data = tall)
```

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

```
summary(m12)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Sex + Semester + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1439
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9880 -0.5266 -0.0270  0.4957  3.8612
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.53969  0.7346
##           RubricInitEDA   0.34918  0.5909   0.47
##           RubricInterpRes 0.20149  0.4489   0.26 0.77
```

```
##           RubricRsrchQ    0.17738  0.4212   0.58 0.47 0.75
##           RubricSelMeth   0.07178  0.2679   0.40 0.63 0.79 0.48
##           RubricTxtOrg    0.26916  0.5188   0.35 0.63 0.76 0.58 0.69
##           RubricVisOrg    0.28013  0.5293   0.36 0.76 0.72 0.57 0.50 0.79
##  Residual                 0.19185  0.4380
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      2.65755    0.39933   6.655
## RubricInitEDA    0.54260    0.09490   5.717
## RubricInterpRes  0.58027    0.09992   5.808
## RubricRsrchQ     0.45482    0.08673   5.244
## RubricSelMeth    0.15818    0.09277   1.705
## RubricTxtOrg     0.68682    0.09847   6.975
## RubricVisOrg     0.52306    0.09820   5.327
## SexF            -0.71191    0.39730  -1.792
## SexM            -0.69683    0.39648  -1.758
## SemesterS19     -0.16809    0.09128  -1.842
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO SexF   SexM
## RubrcIntEDA -0.148
## RbrcIntrpRs -0.174  0.736
## RubrcRsrchQ -0.166  0.589  0.754
## RubricSlMth -0.184  0.664  0.772  0.689
## RubrcTxtOrg -0.163  0.674  0.758  0.679  0.724
## RubricVsOrg -0.160  0.719  0.743  0.670  0.679  0.754
## SexF        -0.966  0.000  0.000 -0.002 -0.005  0.000 -0.001
## SexM        -0.968  0.000  0.000 -0.002 -0.005  0.000 -0.001  0.977
## SemesterS19 -0.001 -0.002 -0.001  0.002  0.006 -0.001 -0.001 -0.099 -0.035
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
m13 = lmer(Rating ~ 1 + Rubric*Sex+(0+Rubric | Artifact),data = tall)
summary(m13)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric * Sex + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1438.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9524 -0.5418 -0.0174  0.4892  3.8757
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.54406  0.7376
##           RubricInitEDA   0.34794  0.5899   0.46
##           RubricInterpRes 0.20035  0.4476   0.25 0.75
##           RubricRsrchQ    0.15988  0.3998   0.60 0.45 0.72
##           RubricSelMeth   0.08183  0.2861   0.37 0.61 0.82 0.50
##           RubricTxtOrg    0.27996  0.5291   0.34 0.63 0.72 0.58 0.74
```

```
##           RubricVisOrg    0.28284  0.5318   0.39 0.77 0.73 0.54 0.59 0.78
##  Residual                 0.19304  0.4394
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)             3.000e+00  8.585e-01   3.494
## RubricInitEDA          -7.670e-13  9.387e-01   0.000
## RubricInterpRes        -3.077e-13  9.818e-01   0.000
## RubricRsrchQ           -9.588e-13  8.576e-01   0.000
## RubricSelMeth          -1.555e-12  9.250e-01   0.000
## RubricTxtOrg           -2.531e-13  9.711e-01   0.000
## RubricVisOrg            3.245e-13  9.537e-01   0.000
## SexF                   -1.164e+00  8.670e-01  -1.343
## SexM                   -1.028e+00  8.689e-01  -1.183
## RubricInitEDA:SexF      5.818e-01  9.476e-01   0.614
## RubricInterpRes:SexF    6.608e-01  9.912e-01   0.667
## RubricRsrchQ:SexF       5.553e-01  8.656e-01   0.642
## RubricSelMeth:SexF      1.355e-01  9.336e-01   0.145
## RubricTxtOrg:SexF       7.573e-01  9.804e-01   0.772
## RubricVisOrg:SexF       6.507e-01  9.628e-01   0.676
## RubricInitEDA:SexM      5.051e-01  9.496e-01   0.532
## RubricInterpRes:SexM    4.964e-01  9.933e-01   0.500
## RubricRsrchQ:SexM       3.383e-01  8.673e-01   0.390
## RubricSelMeth:SexM      1.936e-01  9.356e-01   0.207
## RubricTxtOrg:SexM       6.143e-01  9.825e-01   0.625
## RubricVisOrg:SexM       3.816e-01  9.648e-01   0.395
```

```
##
## Correlation matrix not shown by default, as p = 21 > 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| = 0.00202467 (tol = 0.002, component 1)
```

```r
m14 = lmer(Rating ~ 1 + Rubric*Rater+(0+Rubric | Artifact),data = tall)
```

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

```r
summary(m14)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric * Rater + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1432.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9035 -0.5196 -0.0339  0.4855  3.5712
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.50452  0.7103
##           RubricInitEDA   0.34526  0.5876   0.45
```

```
##           RubricInterpRes 0.15386  0.3922   0.37 0.81
##           RubricRsrchQ    0.16401  0.4050   0.64 0.43 0.70
##           RubricSelMeth   0.08735  0.2956   0.47 0.62 0.77 0.41
##           RubricTxtOrg    0.26331  0.5131   0.43 0.64 0.69 0.55 0.67
##           RubricVisOrg    0.26139  0.5113   0.36 0.72 0.70 0.52 0.46 0.79
##  Residual                 0.18650  0.4319
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)                1.7108     0.1150  14.870
## RubricInitEDA              0.7444     0.1365   5.455
## RubricInterpRes            1.0124     0.1344   7.530
## RubricRsrchQ               0.7468     0.1239   6.028
## RubricSelMeth              0.4244     0.1297   3.272
## RubricTxtOrg               1.0486     0.1352   7.754
## RubricVisOrg               0.6826     0.1390   4.909
## Rater2                     0.3681     0.1332   2.764
## Rater3                     0.2270     0.1327   1.711
## RubricInitEDA:Rater2      -0.3074     0.1726  -1.781
## RubricInterpRes:Rater2    -0.5370     0.1701  -3.157
## RubricRsrchQ:Rater2       -0.4994     0.1613  -3.097
## RubricSelMeth:Rater2      -0.3997     0.1649  -2.424
## RubricTxtOrg:Rater2       -0.5854     0.1715  -3.413
## RubricVisOrg:Rater2       -0.1474     0.1744  -0.845
## RubricInitEDA:Rater3      -0.3058     0.1721  -1.777
## RubricInterpRes:Rater3    -0.7537     0.1697  -4.441
## RubricRsrchQ:Rater3       -0.3774     0.1607  -2.348
## RubricSelMeth:Rater3      -0.3997     0.1644  -2.431
## RubricTxtOrg:Rater3       -0.4980     0.1711  -2.911
## RubricVisOrg:Rater3       -0.3390     0.1740  -1.948

##
## Correlation matrix not shown by default, as p = 21 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

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

```
m15=lmer(Rating ~ 1 + Rater * Semester+(0+Rubric | Artifact),data = tall)
```

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

```
summary(m15)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater * Semester + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1481.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1778 -0.4961 -0.0697  0.5356  3.6921
##
```

```
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.66360  0.8146
##           RubricInitEDA   0.38120  0.6174    0.28
##           RubricInterpRes 0.22042  0.4695   -0.01  0.78
##           RubricRsrchQ    0.18333  0.4282    0.42  0.51  0.73
##           RubricSelMeth   0.08477  0.2912    0.61  0.34  0.31  0.24
##           RubricTxtOrg    0.36824  0.6068    0.02  0.67  0.78  0.61  0.14
##           RubricVisOrg    0.28475  0.5336    0.18  0.77  0.74  0.57  0.19  0.77
##  Residual                 0.19240  0.4386
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           2.34765    0.05932  39.576
## Rater2                0.01343    0.06408   0.210
## Rater3               -0.13239    0.06408  -2.066
## SemesterS19          -0.12782    0.10960  -1.166
## Rater2:SemesterS19 -0.04734    0.12628  -0.375
## Rater3:SemesterS19 -0.13494    0.12627  -1.069
##
## Correlation of Fixed Effects:
##             (Intr) Rater2 Rater3 SmsS19 R2:SS1
## Rater2      -0.550
## Rater3      -0.550  0.504
## SemesterS19 -0.541  0.298  0.298
## Rtr2:SmsS19  0.279 -0.507 -0.256 -0.561
## Rtr3:SmsS19  0.279 -0.256 -0.507 -0.561  0.493
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```r
m16=lmer(Rating ~ 1 + Rubric * Semester+(0+Rubric | Artifact),data = tall)
summary(m16)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric * Semester + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1440.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0001 -0.5262 -0.0328  0.5276  3.8870
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.55484  0.7449
##           RubricInitEDA   0.35056  0.5921   0.47
##           RubricInterpRes 0.20118  0.4485   0.26 0.75
##           RubricRsrchQ    0.16426  0.4053   0.60 0.45 0.74
##           RubricSelMeth   0.07455  0.2730   0.42 0.66 0.83 0.56
##           RubricTxtOrg    0.27330  0.5228   0.34 0.63 0.72 0.61 0.69
##           RubricVisOrg    0.28289  0.5319   0.37 0.75 0.72 0.59 0.49 0.77
##  Residual                 0.18975  0.4356
## Number of obs: 817, groups:  Artifact, 91
```

```
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                  1.94010    0.10687  18.153
## RubricInitEDA                0.52190    0.11371   4.590
## RubricInterpRes              0.57399    0.12002   4.782
## RubricRsrchQ                 0.39251    0.10292   3.814
## RubricSelMeth                0.23268    0.11061   2.104
## RubricTxtOrg                 0.71025    0.11859   5.989
## RubricVisOrg                 0.55017    0.11782   4.669
## SemesterS19                 -0.10405    0.19529  -0.533
## RubricInitEDA:SemesterS19    0.07039    0.20827   0.338
## RubricInterpRes:SemesterS19  0.02608    0.21953   0.119
## RubricRsrchQ:SemesterS19     0.20842    0.18928   1.101
## RubricSelMeth:SemesterS19   -0.25390    0.20292  -1.251
## RubricTxtOrg:SemesterS19    -0.07776    0.21692  -0.358
## RubricVisOrg:SemesterS19    -0.08813    0.21525  -0.409
##
## Correlation matrix not shown by default, as p = 14 > 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| = 0.00376018 (tol = 0.002, component 1)
```

```
m17=lmer(Rating ~ 1 + Rubric + Rater+Semester+Rater*Rubric+(0+Rubric | Artifact),data = tall)
```

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

```
summary(m17)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Semester + Rater * Rubric + (0 +
##     Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1430.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9358 -0.5143 -0.0409  0.4895  3.5799
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.50835  0.7130
##           RubricInitEDA   0.35401  0.5950   0.46
##           RubricInterpRes 0.15703  0.3963   0.38 0.82
##           RubricRsrchQ    0.18268  0.4274   0.64 0.45 0.73
##           RubricSelMeth   0.07262  0.2695   0.45 0.62 0.76 0.41
##           RubricTxtOrg    0.25908  0.5090   0.43 0.64 0.68 0.56 0.65
##           RubricVisOrg    0.25611  0.5061   0.36 0.72 0.69 0.53 0.41 0.78
##  Residual                 0.18484  0.4299
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
```

```
##                         Estimate Std. Error t value
## (Intercept)              1.77051    0.11793  15.013
## RubricInitEDA            0.74459    0.13626   5.465
## RubricInterpRes          1.01239    0.13425   7.541
## RubricRsrchQ             0.74688    0.12377   6.035
## RubricSelMeth            0.42339    0.12972   3.264
## RubricTxtOrg             1.04782    0.13506   7.758
## RubricVisOrg             0.68169    0.13886   4.909
## Rater2                   0.36516    0.13281   2.750
## Rater3                   0.22497    0.13235   1.700
## SemesterS19             -0.18905    0.08384  -2.255
## RubricInitEDA:Rater2    -0.30780    0.17212  -1.788
## RubricInterpRes:Rater2 -0.53771    0.16967  -3.169
## RubricRsrchQ:Rater2     -0.50228    0.16112  -3.117
## RubricSelMeth:Rater2    -0.39806    0.16417  -2.425
## RubricTxtOrg:Rater2     -0.58388    0.17104  -3.414
## RubricVisOrg:Rater2     -0.14567    0.17393  -0.838
## RubricInitEDA:Rater3    -0.30463    0.17167  -1.775
## RubricInterpRes:Rater3 -0.75345    0.16928  -4.451
## RubricRsrchQ:Rater3     -0.37509    0.16057  -2.336
## RubricSelMeth:Rater3    -0.40351    0.16374  -2.464
## RubricTxtOrg:Rater3     -0.49780    0.17061  -2.918
## RubricVisOrg:Rater3     -0.33878    0.17354  -1.952
##
## 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
```

```
m18=lmer(Rating ~ 1 + Rubric + Rater+Semester+Rater*Semester+(0+Rubric | Artifact),data = tall)
```

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

```
summary(m18)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Semester + Rater * Semester + (0 +
##     Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1439.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1044 -0.5167 -0.0217  0.5249  3.7936
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.5597   0.7481
##           RubricInitEDA   0.3476   0.5896   0.48
##           RubricInterpRes 0.1677   0.4095   0.25 0.75
##           RubricRsrchQ    0.1730   0.4160   0.60 0.46 0.72
##           RubricSelMeth   0.0669   0.2586   0.44 0.62 0.75 0.45
```

```
##           RubricTxtOrg    0.2484   0.4984   0.34 0.62 0.70 0.57 0.66
##           RubricVisOrg    0.2555   0.5055   0.36 0.74 0.69 0.53 0.43 0.76
##  Residual                 0.1906   0.4366
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           2.00753    0.10063  19.950
## RubricInitEDA         0.54188    0.09490   5.710
## RubricInterpRes       0.58020    0.10002   5.801
## RubricRsrchQ          0.45346    0.08667   5.232
## RubricSelMeth         0.15664    0.09274   1.689
## RubricTxtOrg          0.68615    0.09874   6.949
## RubricVisOrg          0.52263    0.09822   5.321
## Rater2                0.01147    0.06390   0.180
## Rater3               -0.13562    0.06390  -2.122
## SemesterS19          -0.12691    0.10974  -1.156
## Rater2:SemesterS19 -0.04922    0.12599  -0.391
## Rater3:SemesterS19 -0.13451    0.12598  -1.068
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO Rater2 Rater3
## RubrcIntEDA -0.599
## RbrcIntrpRs -0.720  0.734
## RubrcRsrchQ -0.684  0.588  0.756
## RubricSlMth -0.764  0.663  0.778  0.688
## RubrcTxtOrg -0.669  0.676  0.752  0.684  0.729
## RubricVsOrg -0.662  0.716  0.745  0.668  0.682  0.752
## Rater2      -0.323  0.000  0.000  0.000  0.000  0.000 -0.001
## Rater3      -0.323  0.000  0.000  0.000  0.000  0.000 -0.001  0.504
## SemesterS19 -0.322 -0.001  0.000  0.002  0.005  0.000  0.000  0.296  0.296
## Rtr2:SmsS19  0.165 -0.001 -0.001 -0.001 -0.002 -0.001 -0.001 -0.507 -0.255
## Rtr3:SmsS19  0.164  0.000  0.000  0.000 -0.001  0.000  0.000 -0.255 -0.507
##             SmsS19 R2:SS1
## RubrcIntEDA
## RbrcIntrpRs
## RubrcRsrchQ
## RubricSlMth
## RubrcTxtOrg
## RubricVsOrg
## Rater2
## Rater3
## SemesterS19
## Rtr2:SmsS19 -0.559
## Rtr3:SmsS19 -0.559  0.493
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
m19=lmer(Rating ~ 1 + Rubric + Rater+Semester+Semester*Rubric+(0+Rubric | Artifact),data=tall)
summary(m19)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Semester + Semester * Rubric +
##     (0 + Rubric | Artifact)
##    Data: tall
```

```
##
## REML criterion at convergence: 1436.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1412 -0.5223 -0.0391  0.5445  3.7772
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.56851  0.7540
##           RubricInitEDA   0.35082  0.5923   0.47
##           RubricInterpRes 0.17159  0.4142   0.24 0.74
##           RubricRsrchQ    0.15668  0.3958   0.62 0.43 0.71
##           RubricSelMeth   0.06959  0.2638   0.44 0.65 0.80 0.52
##           RubricTxtOrg    0.26019  0.5101   0.34 0.62 0.69 0.58 0.66
##           RubricVisOrg    0.26379  0.5136   0.36 0.74 0.69 0.56 0.43 0.75
##  Residual                 0.18786  0.4334
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                 1.997781   0.112468  17.763
## RubricInitEDA               0.521859   0.113698   4.590
## RubricInterpRes             0.573297   0.120045   4.776
## RubricRsrchQ                0.392100   0.102918   3.810
## RubricSelMeth               0.232367   0.110677   2.099
## RubricTxtOrg                0.709756   0.118618   5.984
## RubricVisOrg                0.550031   0.117885   4.666
## Rater2                     -0.001354   0.054870  -0.025
## Rater3                     -0.168341   0.054869  -3.068
## SemesterS19                -0.106703   0.196877  -0.542
## RubricInitEDA:SemesterS19   0.070051   0.208228   0.336
## RubricInterpRes:SemesterS19  0.026036   0.219571   0.119
## RubricRsrchQ:SemesterS19    0.208839   0.189252   1.103
## RubricSelMeth:SemesterS19  -0.253631   0.203031  -1.249
## RubricTxtOrg:SemesterS19   -0.077956   0.216962  -0.359
## RubricVisOrg:SemesterS19   -0.088057   0.215348  -0.409
```

```
##
## Correlation matrix not shown by default, as p = 16 > 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| = 0.0070201 (tol = 0.002, component 1)
```

```r
m20=lmer(Rating ~ 1 + Rubric + Rater+Semester+Rater*Rubric+Rater*Semester+Semester*Rubric+(0+Rubric |
summary(m20)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Semester + Rater * Rubric + Rater *
##     Semester + Semester * Rubric + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1436.3
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9004 -0.5233 -0.0569  0.5022  3.5936
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.50976  0.7140
##           RubricInitEDA   0.35049  0.5920   0.45
##           RubricInterpRes 0.15264  0.3907   0.36 0.81
##           RubricRsrchQ    0.16766  0.4095   0.65 0.43 0.72
##           RubricSelMeth   0.06979  0.2642   0.47 0.65 0.80 0.49
##           RubricTxtOrg    0.25859  0.5085   0.42 0.64 0.67 0.58 0.64
##           RubricVisOrg    0.25524  0.5052   0.35 0.72 0.69 0.55 0.39 0.77
##  Residual                 0.18524  0.4304
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                              Estimate Std. Error t value
## (Intercept)                  1.725226   0.130479  13.222
## RubricInitEDA                0.727960   0.151044   4.820
## RubricInterpRes              1.009669   0.148889   6.781
## RubricRsrchQ                 0.684446   0.135484   5.052
## RubricSelMeth                0.507216   0.143096   3.545
## RubricTxtOrg                 1.078154   0.149730   7.201
## RubricVisOrg                 0.711894   0.154245   4.615
## Rater2                       0.375824   0.136605   2.751
## Rater3                       0.255422   0.136145   1.876
## SemesterS19                 -0.032192   0.201163  -0.160
## RubricInitEDA:Rater2        -0.303909   0.172515  -1.762
## RubricInterpRes:Rater2      -0.533383   0.170170  -3.134
## RubricRsrchQ:Rater2         -0.491185   0.160696  -3.057
## RubricSelMeth:Rater2        -0.402902   0.164295  -2.452
## RubricTxtOrg:Rater2         -0.584148   0.171597  -3.404
## RubricVisOrg:Rater2         -0.146279   0.174442  -0.839
## RubricInitEDA:Rater3        -0.302639   0.172021  -1.759
## RubricInterpRes:Rater3      -0.749842   0.169751  -4.417
## RubricRsrchQ:Rater3         -0.366243   0.160093  -2.288
## RubricSelMeth:Rater3        -0.404292   0.163823  -2.468
## RubricTxtOrg:Rater3         -0.497374   0.171131  -2.906
## RubricVisOrg:Rater3         -0.337815   0.174019  -1.941
## Rater2:SemesterS19          -0.045130   0.124735  -0.362
## Rater3:SemesterS19          -0.128044   0.124729  -1.027
## RubricInitEDA:SemesterS19    0.052734   0.205875   0.256
## RubricInterpRes:SemesterS19  0.002847   0.202999   0.014
## RubricRsrchQ:SemesterS19     0.190731   0.179944   1.060
## RubricSelMeth:SemesterS19   -0.273631   0.194815  -1.405
## RubricTxtOrg:SemesterS19    -0.098918   0.203595  -0.486
## RubricVisOrg:SemesterS19    -0.099740   0.210234  -0.474

##
## 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| = 0.00390185 (tol = 0.002, component 1)
```

performing anova analysis by group we have:

```
anova(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12)
```

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

```
## Data: tall
## Models:
## m1: Rating ~ 1 + Rater + (0 + Rubric | Artifact)
## m3: Rating ~ 1 + Sex + (0 + Rubric | Artifact)
## m5: Rating ~ 1 + Rater + Sex + (0 + Rubric | Artifact)
## m8: Rating ~ 1 + Rater + Sex + Repeated + (0 + Rubric | Artifact)
## m11: Rating ~ 1 + Rater + Sex + Semester + (0 + Rubric | Artifact)
## m2: Rating ~ 1 + Rubric + (0 + Rubric | Artifact)
## m4: Rating ~ 1 + Rubric + Rater + (0 + Rubric | Artifact)
## m6: Rating ~ 1 + Rubric + Sex + (0 + Rubric | Artifact)
## m7: Rating ~ 1 + Rubric + Rater + Repeated + (0 + Rubric | Artifact)
## m9: Rating ~ 1 + Rubric + Sex + Repeated + (0 + Rubric | Artifact)
## m10: Rating ~ 1 + Rater + Semester + Rubric + (0 + Rubric | Artifact)
## m12: Rating ~ 1 + Rubric + Sex + Semester + (0 + Rubric | Artifact)
##     npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m1    32 1529.9 1680.5 -732.94   1465.9
## m3    32 1536.9 1687.5 -736.43   1472.9  0.0000  0
## m5    34 1528.2 1688.1 -730.07   1460.2 12.7169  2   0.001732 **
## m8    35 1529.3 1694.0 -729.64   1459.3  0.8769  1   0.349054
## m11   35 1526.7 1691.3 -728.32   1456.7  2.6261  0
## m2    36 1485.0 1654.4 -706.51   1413.0 43.6218  1  3.984e-11 ***
## m4    38 1477.5 1656.3 -700.73   1401.5 11.5618  2   0.003086 **
## m6    38 1485.0 1663.8 -704.49   1409.0  0.0000  0
## m7    39 1479.1 1662.6 -700.53   1401.1  7.9268  1   0.004871 **
## m9    39 1486.6 1670.2 -704.32   1408.6  0.0000  0
## m10   39 1475.2 1658.7 -698.58   1397.2 11.4734  0
## m12   39 1484.1 1667.6 -703.05   1406.1  0.0000  0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the ANOVA table above, we find that model2, which has the lowest AIC and BIC values, also has lowest p-value. So we find the the term Rubricvery useful, that with different Rubric, different Rater tend to give different Ratings for a specfic artifect.

```
#Final model selected by anova:
m2 = lmer(Rating ~ 1 + Rubric + (0+Rubric | Artifact),data = tall)
```

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

```
summary(m2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1439.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9610 -0.5301 -0.0116  0.5062  3.8765
```

```
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.54648  0.7392
##           RubricInitEDA   0.34214  0.5849   0.47
##           RubricInterpRes 0.19783  0.4448   0.26 0.76
##           RubricRsrchQ    0.16164  0.4020   0.59 0.45 0.72
##           RubricSelMeth   0.08874  0.2979   0.43 0.63 0.79 0.48
##           RubricTxtOrg    0.27144  0.5210   0.35 0.64 0.74 0.59 0.71
##           RubricVisOrg    0.28462  0.5335   0.37 0.75 0.73 0.56 0.53 0.79
##  Residual                 0.19251  0.4388
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      1.90899    0.08907  21.432
## RubricInitEDA    0.54139    0.09491   5.704
## RubricInterpRes  0.58064    0.09999   5.807
## RubricRsrchQ     0.45341    0.08680   5.224
## RubricSelMeth    0.15830    0.09228   1.715
## RubricTxtOrg     0.68645    0.09870   6.955
## RubricVisOrg     0.52298    0.09813   5.329
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO
## RubrcIntEDA -0.673
## RbrcIntrpRs -0.787  0.734
## RubrcRsrchQ -0.771  0.589  0.751
## RubricSlMth -0.836  0.667  0.775  0.690
## RubrcTxtOrg -0.734  0.677  0.752  0.683  0.729
## RubricVsOrg -0.722  0.718  0.744  0.667  0.681  0.752
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
anova(m13,m14,m15,m16,m17,m18,m19,m20)
```

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

```
## Data: tall
## Models:
## m15: Rating ~ 1 + Rater * Semester + (0 + Rubric | Artifact)
## m18: Rating ~ 1 + Rubric + Rater + Semester + Rater * Semester + (0 + Rubric | Artifact)
## m16: Rating ~ 1 + Rubric * Semester + (0 + Rubric | Artifact)
## m19: Rating ~ 1 + Rubric + Rater + Semester + Semester * Rubric + (0 + Rubric | Artifact)
## m13: Rating ~ 1 + Rubric * Sex + (0 + Rubric | Artifact)
## m14: Rating ~ 1 + Rubric * Rater + (0 + Rubric | Artifact)
## m17: Rating ~ 1 + Rubric + Rater + Semester + Rater * Rubric + (0 + Rubric | Artifact)
## m20: Rating ~ 1 + Rubric + Rater + Semester + Rater * Rubric + Rater * Semester + Semester * Rubri
##     npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m15   35 1530.3 1695.0 -730.16   1460.3
## m18   41 1478.0 1670.9 -697.99   1396.0 64.3254  6  5.924e-12 ***
## m16   43 1483.5 1685.8 -698.74   1397.5  0.0000  2   1.000000
## m19   45 1475.5 1687.3 -692.76   1385.5 11.9714  2   0.002514 **
## m13   50 1501.0 1736.3 -700.52   1401.0  0.0000  5   1.000000
## m14   50 1468.1 1703.4 -684.06   1368.1 32.9164  0
```

```
## m17   51 1465.5 1705.5 -681.76   1363.5  4.6115  1   0.031758 *
## m20   59 1468.9 1746.5 -675.44   1350.9 12.6426  8   0.124754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the ANOVA table above, we find that model18, which does not have the lowest AIC and BIC values, has lowest p-value. So we find the the interaction term Rater*Semester very useful, that with different Semesters, different Rater tend to give different Ratings for a specfic artifect.

```
summary(m18)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rubric + Rater + Semester + Rater * Semester + (0 +
##     Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1439.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1044 -0.5167 -0.0217  0.5249  3.7936
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.5597   0.7481
##           RubricInitEDA   0.3476   0.5896   0.48
##           RubricInterpRes 0.1677   0.4095   0.25 0.75
##           RubricRsrchQ    0.1730   0.4160   0.60 0.46 0.72
##           RubricSelMeth   0.0669   0.2586   0.44 0.62 0.75 0.45
##           RubricTxtOrg    0.2484   0.4984   0.34 0.62 0.70 0.57 0.66
##           RubricVisOrg    0.2555   0.5055   0.36 0.74 0.69 0.53 0.43 0.76
##  Residual                 0.1906   0.4366
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)         2.00753    0.10063  19.950
## RubricInitEDA       0.54188    0.09490   5.710
## RubricInterpRes     0.58020    0.10002   5.801
## RubricRsrchQ        0.45346    0.08667   5.232
## RubricSelMeth       0.15664    0.09274   1.689
## RubricTxtOrg        0.68615    0.09874   6.949
## RubricVisOrg        0.52263    0.09822   5.321
## Rater2              0.01147    0.06390   0.180
## Rater3             -0.13562    0.06390  -2.122
## SemesterS19        -0.12691    0.10974  -1.156
## Rater2:SemesterS19 -0.04922    0.12599  -0.391
## Rater3:SemesterS19 -0.13451    0.12598  -1.068
##
## Correlation of Fixed Effects:
##             (Intr) RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO RbrcVO Rater2 Rater3
## RubrcIntEDA -0.599
## RbrcIntrpRs -0.720  0.734
## RubrcRsrchQ -0.684  0.588  0.756
```

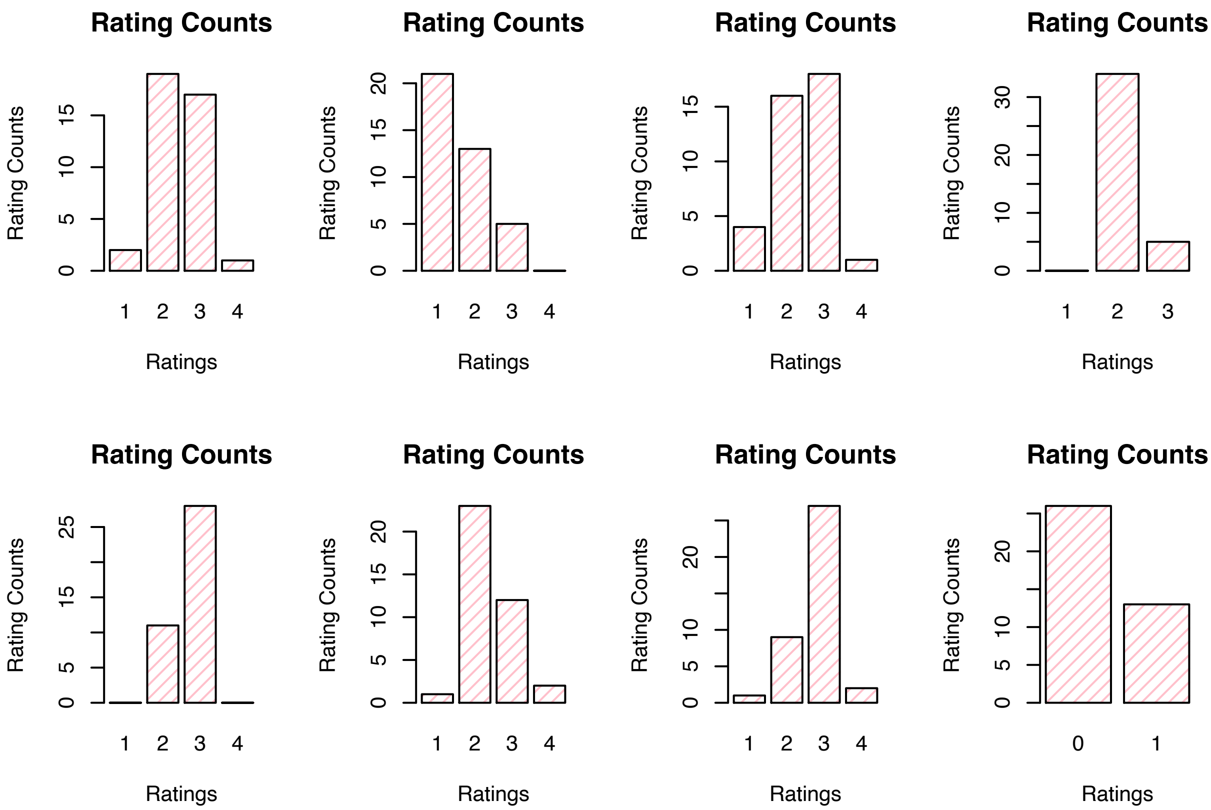
```
## RubricSlMth -0.764  0.663  0.778  0.688
## RubrcTxtOrg -0.669  0.676  0.752  0.684  0.729
## RubricVsOrg -0.662  0.716  0.745  0.668  0.682  0.752
## Rater2      -0.323  0.000  0.000  0.000  0.000  0.000 -0.001
## Rater3      -0.323  0.000  0.000  0.000  0.000  0.000 -0.001  0.504
## SemesterS19 -0.322 -0.001  0.000  0.002  0.005  0.000  0.000  0.296  0.296
## Rtr2:SmsS19  0.165 -0.001 -0.001 -0.001 -0.002 -0.001 -0.001 -0.507 -0.255
## Rtr3:SmsS19  0.164  0.000  0.000  0.000 -0.001  0.000  0.000 -0.255 -0.507
##             SmsS19 R2:SS1
## RubrcIntEDA
## RbrcIntrpRs
## RubrcRsrchQ
## RubricSlMth
## RubrcTxtOrg
## RubricVsOrg
## Rater2
## Rater3
## SemesterS19
## Rtr2:SmsS19 -0.559
## Rtr3:SmsS19 -0.559  0.493
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Choosing model 18 over model 2 is probably a good decision because, we find that different rubric and different rater could actually have strong effect on the rating of each artifact. The interaction terms in our model are indeed interesting, since the interaction terms actually show that for different rubric and different rater, each artifact tend to have different ratings.It is necessary to have interaction terms in our final model.

Research Question 4

An interesting question I found here is the barplots for each rating score groups:

```
#Looking at the barplot of rating score=1 as an example:

par(mfrow=c(2,4))
ratings_1<- ratings %>% filter(ratings$Rater==1)

barplot(table(ratings_1$RsrchQ),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$CritDes),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$InitEDA),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$SelMeth),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",

barplot(table(ratings_1$InterpRes),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink

barplot(table(ratings_1$VisOrg),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$TxtOrg),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink",d

barplot(table(ratings_1$Repeated),main="Rating Counts",xlab="Ratings",ylab="Rating Counts",col="pink"
```



This is significantly different accross rating groups. I was wondering what caused these differences and what are possible affects these distributions might have on our study results. Some distributions are extremely skewed for certain rating groups.

I think it is worth trying to subset the data into three rating score groups and perform analysis on each group of data. With a final model for each group, we can cross-validate the models to find out which one works for the whole population.