Sesame Street Report [Draft]

Armelle, Sara, Ibrohim

2024-02-05

Table of contents

1 1. Project Description

The project involves an observational study designed to assess the impact of "Sesame Street" viewership on children's learning outcomes, particularly in letters, numbers, and forms. The study is in the analysis stage, examining existing data collected from children across five different sites in the United States. Variables observed include, but are not limited to, viewing frequency, setting, encouragement to watch, and pretest scores of vocabulary maturity (Peabody Picture Vocabulary Test). Its objective is to determine the show's effectiveness in educational content delivery and identify areas for improvement. Results will be used to enhance Sesame Street's educational focus and impact, as per the client's request for an upcoming board meeting presentation.

2 1.1 Research Questions

Question 1: How does viewership of Sesame Street influence children's learning outcomes - specifically in terms of letters, numbers, and forms knowledge - as measured by the percent increase in post-test scores compared to pre-test scores? Is this result consistent if we instead consider percent achievable gain (PAG)?

Question 2: Which area, of letters, numbers, and forms, has the most room for improvement? Which should Sesame street focus on improving in the future?

3 1.2 Variables

We considered a variety of variables in our preliminary analysis of the Sesame Street study. The dataset contained several explanatory variables, but we chose to focus on viewing frequency (Viewcat) as our main explanatory variable; and site, sex, age, setting, and encouragement as possible confounding variables. The dataset contained pre- and post-test scores for six different domains, but we were only interested in scores for letters, numbers, and forms. We also defined two possible response variables: percent increase and percent achievable gain (PAG). The first, percent increase, is simply the difference between the posttest score and pretest score, as percentages. The second, PAG, is the child's improvement (post-test score minus pre-test score) divided by total possible improvement (maximum test score minus pretest score). The goal with this measure is to capture improvement while accounting for the fact that advanced students cannot improve as much as those who were not as advanced to begin with. All variables used in our analysis are summarized in the table below. Although there were other variables included in the dataset, we decided that these were the most important ones to answer our client's research questions.

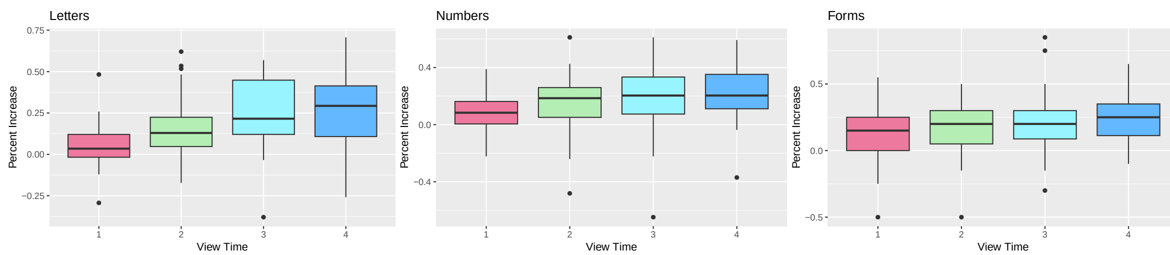
Table 1: Summary of variables used in analysis

Name	Type	Notes
ID	Numerical	Identifying numeric sequence
Site	Categorical	Five different sampling sites (Explanatory)
Sex	Categorical	Male or Female (Explanatory)
Age	Numerical	Age in months (Explanatory)
Viewcat	Categorical	Categorical 1-4 encoding amount of show child watched (Explanatory)
Setting	Categorical	Home or School (Explanatory)
Viewenc	Categorical	Whether or not child was encouraged to view show (Explanatory)
PAG	Numerical	Percent Acheivable Gain (Response)
Percent Increase	Numerical	The difference of post-score percent and pre-score percent (Response)

4 2. Exploratory Data Analysis (EDA)

Our exploratory data analysis of the Sesame Street study revealed promising leads for modeling, as well as concerns about possible data issues and confounders we will have to watch out for as we move forward. As a whole, it appeared that Sesame Street viewership was positively correlated with greater improvement, as measured by PAG and percent increase across all tests (letters, forms, and numbers). The comparative improvement across tests did not appear to have a clear trend measured by both PAG and percent increase.

While PAG is has nice theoretical properties, its use as a response variable in this study may not be practical; it introduced more nonuniform variability in the outcomes. In addition, exploratory analysis revealed that variance may not be equal across all groups (particularly with regards to site), and this could cause possible problems in modeling. More in-depth exploratory analysis is reported below. First, we look at visualizations to answer the two primary research questions; then, we investigate possible confounders.



(a) Letters (b) Numbers (c) Forms

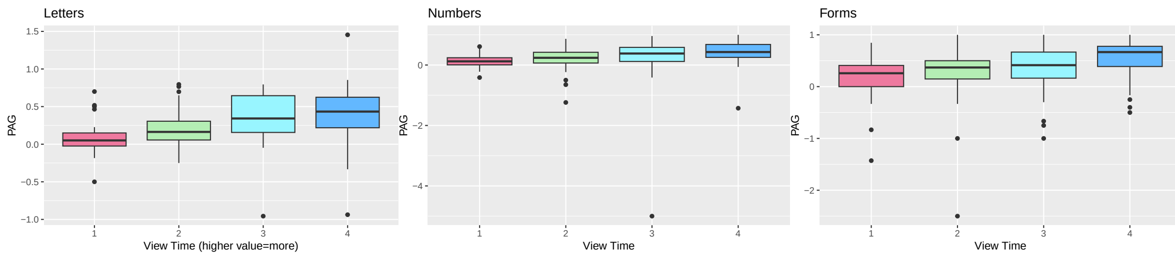
Figure 1: The Relationship Between Viewing Time of Sesame Street Percent Score Increase for Letters, Numbers, and Forms. Note that viewing time is categorical with higher values corresponding to more time viewing Sesame Street

All three plots in Figure 1 appear to indicate a positive linear relationship between viewing time (categorized as levels 1-4) and percent score increase. Although the variability is not completely consistent across viewing levels, overall the data are relatively well-behaved.

Similar to Figure 1, all three plots in Figure 2 seem to indicate a positive relationship between viewing time and percent score increase. However, by comparison to the plots using percent increase as the response variable, the data have a much wider spread with many more outliers. This could pose some problems in future modeling.

The plot in Figure 3 doesn't indicate a clear trend with regards to which subject Sesame Street teaches the most effectively. The highest median is forms, followed by numbers and then letters. However, all three boxplots have significant overlap with each other, which means that statistical analysis will likely not yield usable differences.

Once again, it appears that PAG introduces a lot of additional variance compared to the same plot using percent increase as the response variable. In this case, there is clearly significant



(a) Letters (b) Numbers (c) Forms

Figure 2: The Relationship Between Viewing Time of Sesame Street PAG for Letters, Numbers, and Forms. Note that viewing time is categorical with higher values corresponding to more time viewing Sesame Street

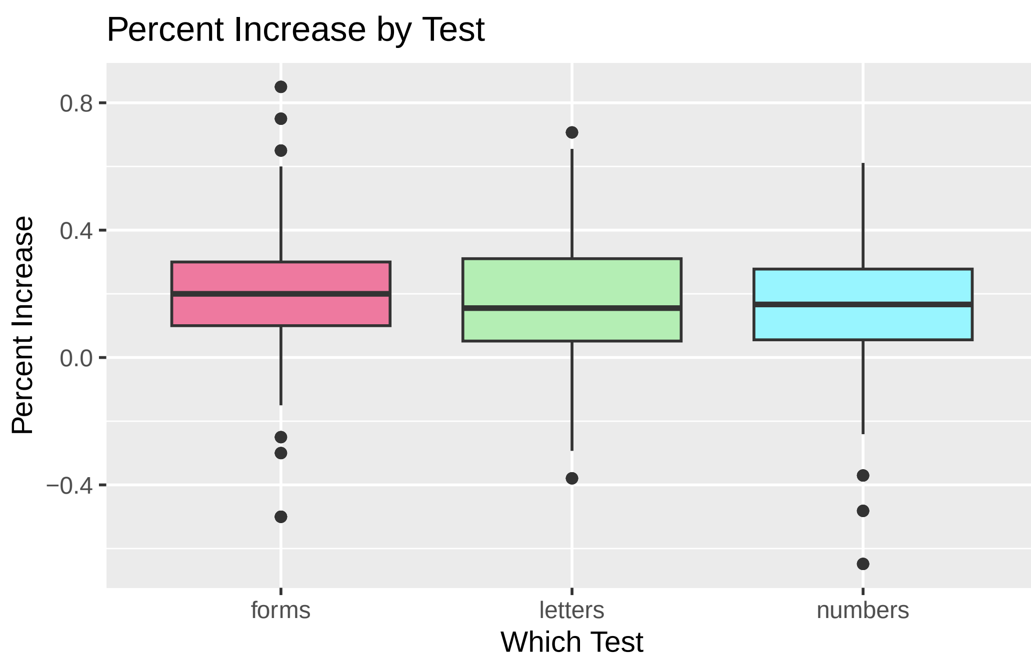


Figure 3: Comparing Percent Increase Across Three Subjects: Letters, Numbers, and Forms

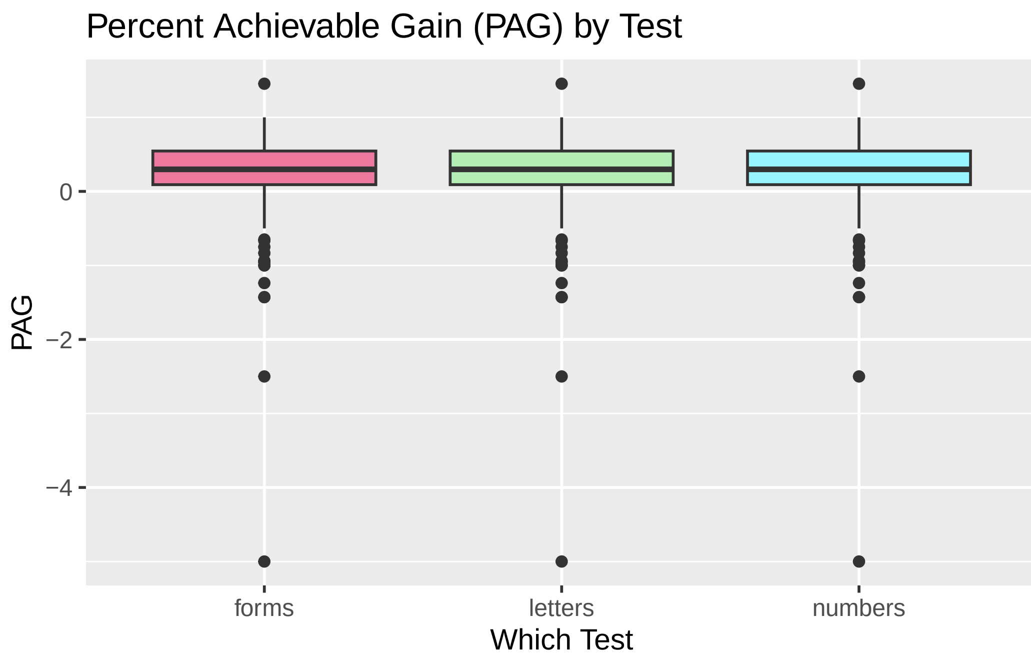
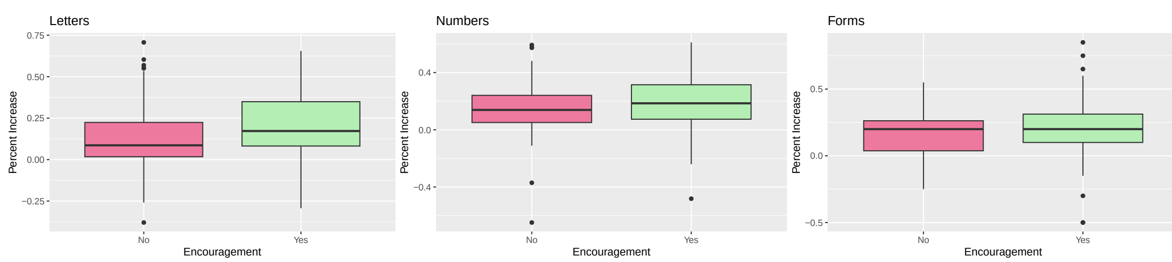


Figure 4: Comparing PAG Across Three Subjects: Letters, Numbers, and Forms

left-skew in the data with lots of outliers. Despite this, the plot in Figure 4 does appear to show similar results as the plot in Figure 3. When put on the same scale, the values appear to be similar around the middle of the boxplot.

For simplicity, all investigation of possible confounders (below), was done with only percent increase as response variable.



(a) Letters (b) Numbers (c) Forms

Figure 5: Comparing Percent Increase Across Three Subjects (Letters, Numbers, and Forms) depending on Encouragement

The plots in Figure 5 have somewhat unclear results. Encouragement may have an impact on percent score increase, particularly for letters, but there is enough overlap of the boxplots in all three plots that more analysis is needed. This may be an important confounder to consider in modeling.

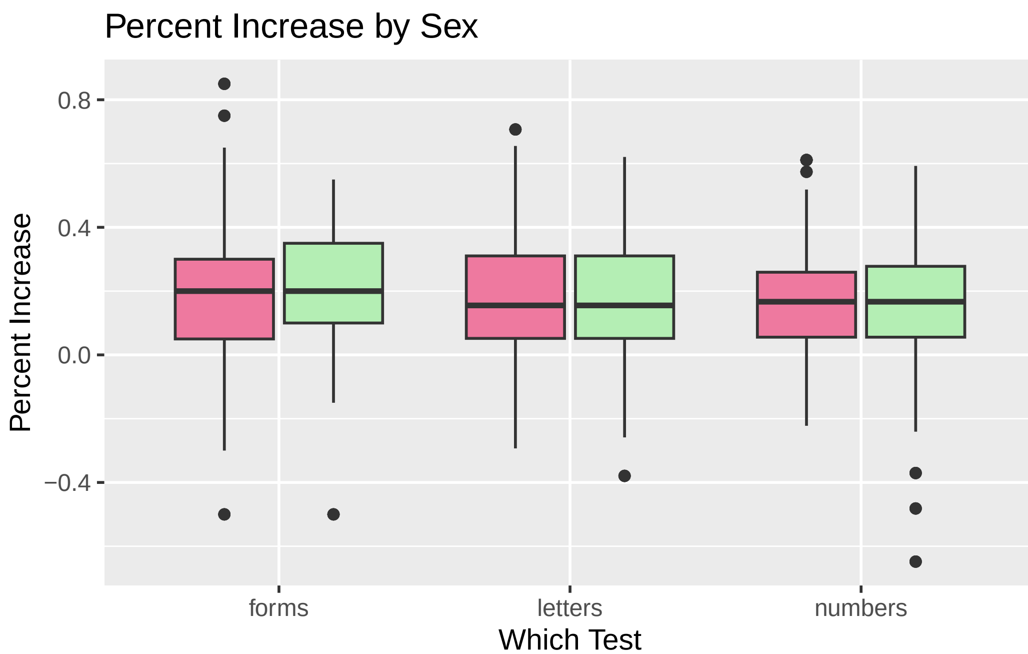


Figure 6: Comparing Percent Increase Across Three Subjects (Letters, Numbers, and Forms) depending on Sex

The plot in Figure 6 seems to show little to no relationship between sex and percent increase in score in any of the tests. It does not appear that sex will be an important confounder to consider in modeling.

The plot in Figure 7 shows some variation in score improvement across the different sites for all three tests. Not only is median percent increase in score different depending on the site, but the variability is not uniform across groups either. This variable seems to be an important confounder, and we will have to look out for issues with the non-uniform varibility.

The plot in Figure 8 seems to show little to no relationship between age and percent increase in score in any of the tests. The points appear to be scattered more or less at random. It does not appear that age will be an important confounder to consider in modeling.

The plot in Figure 9 seems to show little to no relationship between setting and percent increase in score in any of the tests. It does not appear that setting will be an important confounder to consider in modeling.

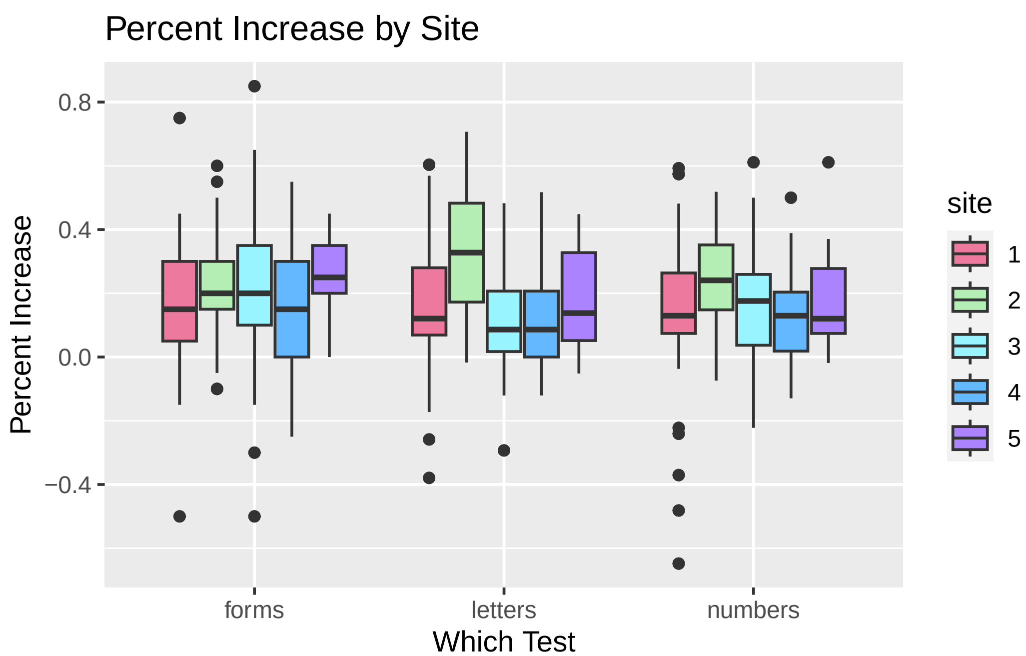


Figure 7: Comparing Percent Increase Across Three Subjects (Letters, Numbers, and Forms) depending on Site

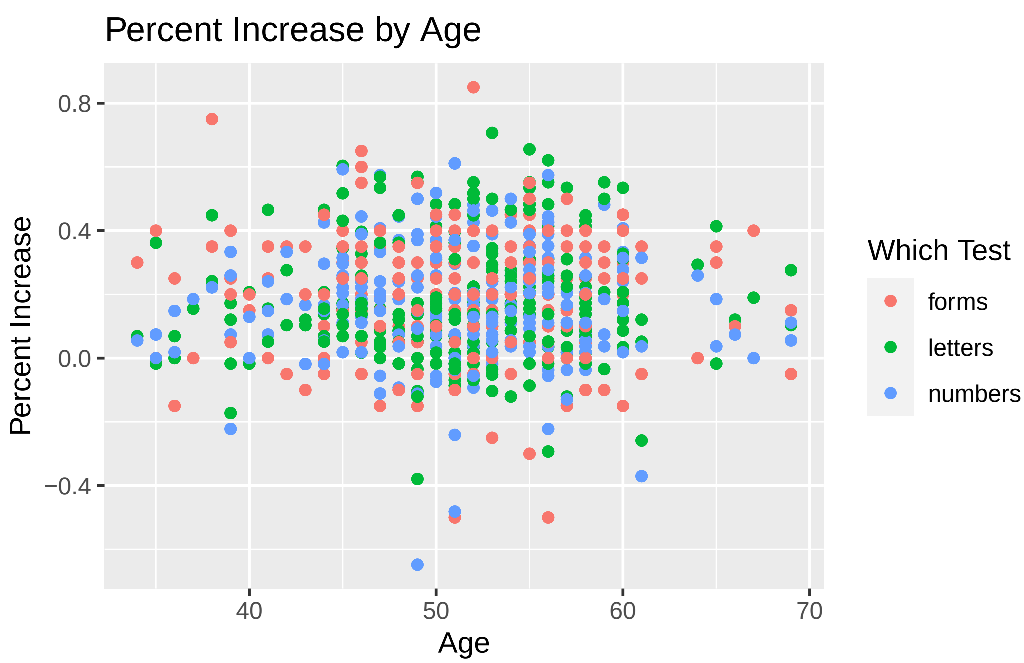


Figure 8: Comparing Percent Increase Across Three Subjects (Letters, Numbers, and Forms) depending on Age

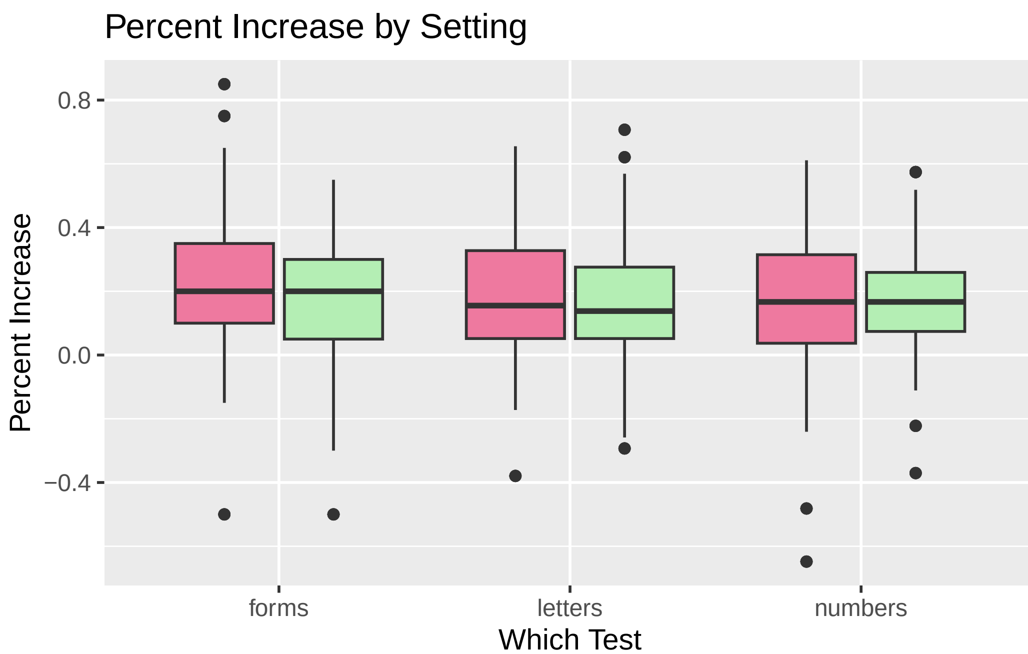


Figure 9: Comparing Percent Increase Across Three Subjects (Letters, Numbers, and Forms) depending on Setting

5 3. Statistical Analysis

To address the first research question, we fit three multiple linear regression models: one for each subject. The models predicted improvement using prescore, age, viewcat, and site. The models were all of the form:

$$\text{SubjectImprovement} = \beta_0 + \beta_1 \text{Prescore} + \beta_2 \text{Age} + \beta_{3 \ldots 5} I_{Viewcat} + \beta_{6 \ldots 9} I_{Site} + \epsilon,$$

$$\epsilon \sim N(0, \sigma).$$

In this model, β_1 and β_2 are attached to the two quantitative variables, and they signify the change in predicted improvement given a unit increase in prescore and age, respectively. The other β's (3 through 9) are attached to the two categorical variables, viewcat and site. The interpretation of β_3 is the change in predicted improvement for viewcat 2 when compared to viewcat 1 (the baseline), and this interpretation extends through to β_5. Similarly, the interpretation of β_6 is the change in predicted improvement for site 2 when compared to site 1 (the baseline), and this interpretation extends through to β_9. Finally, $\beta 0$ is the intercept, and its interpretation is not important in this context. A (type II) ANOVA table for each model is included in Table 2, Table 3, and Table 4.

Table 2: ANOVA table for multiple linear regression model to predict improvement in letters using viewcat, prescore, age, and site

	Sum Sq	Df	F value	Pr(>F)
site	1.3563322	4	15.290852	0.0000000
age	0.1705559	1	7.691167	0.0060044
viewcat	1.4749363	3	22.170609	0.0000000
prelet	0.5921747	1	26.703946	0.0000005
Residuals	5.1003764	230	NA	NA

In this table the values that we are most interested in are the p-values on the right-hand column. Although some values appear to be zero, they are not, but are so small that they were rounded to zero. We will consider any p-value less than 0.05 to be "significant," meaning we can infer that what we observe is a product of the true process and not random chance. For the quantitative variables of age and prescore, the small p-values indicate that each of these variables is important for predicting improvement. For the categorical variables of viewcat and site, the small p-values indicate that at least one of the viewing categories and at least one of the sites resulted in a different improvement than another. The same interpretations can be applied to the two ANOVA tables below, which are analogous to this one but for the forms and numbers models.

Table 3: ANOVA table for multiple linear regression model to predict improvement in forms using viewcat, prescore, age, and site

	Sum Sq	Df	F value	Pr(>F)
site	0.3762414	4	4.141300	0.0029334
age	0.1730415	1	7.618692	0.0062423
viewcat	0.9975679	3	14.640345	0.0000000
preform	2.6819266	1	118.080169	0.0000000
Residuals	5.2239349	230	NA	NA

Table 4: ANOVA table for multiple linear regression model to predict improvement in numbers using viewcat, prescore, age, and site

	Sum Sq	Df	F value	Pr(>F)
site	0.4052706	4	3.943926	0.0040742
age	0.1522685	1	5.927256	0.0156704
viewcat	0.7486290	3	9.713796	0.0000046
prenumb	0.9852310	1	38.351439	0.0000000
Residuals	5.9085953	230	NA	NA

In Figure 10, we can observe the confidence intervals for the mean improvement in letters stratified by category after accounting for pretest score, age, and site. In this plot, if the red arrows for two confidence intervals don't overlap, than we can infer that those two categories are statistically different from each other. In this case, we can see that viewing category 4 was associated with greater improvements than categories 1 and 2, but similar amounts of improvement as category 3 given a particular site, pre-score, and age. Similarly, given a particular site, pre-score, and age, viewing category 3 was associated with greater improvements than categories 1 and 2 and viewing category 2 was associated with greater improvements than category 1.

In Figure 11, we can observe the confidence intervals for the mean improvement in forms stratified by category after accounting for pretest score, age, and site. This plot can be interpreted in the same way as the one for letters, and we can observe similar trends.

Likewise, in Figure 12, we can observe the confidence intervals for the mean improvement in numbers stratified by category after accounting for pretest score, age, and site. Once again, the same interpretations can be applied and the trend is similar to the other two.

To address the second research question, we fit a mixed-effects linear model predicting improvement with subject, age, viewcat, site, and an interaction term between subject and viewcat while accounting for the random effects of the individual child (in the model, we call the

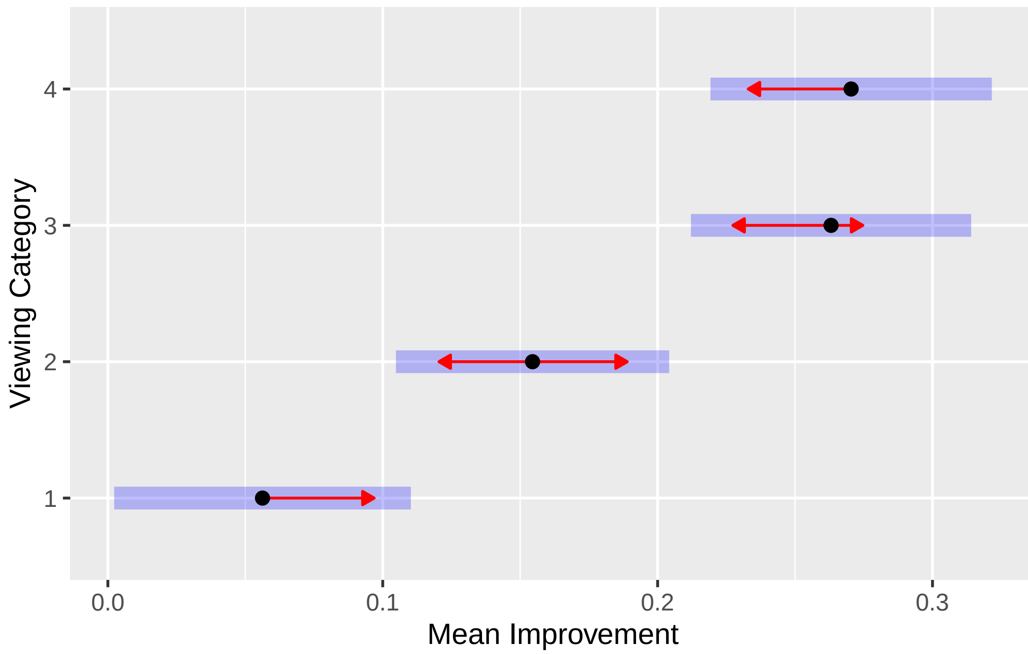


Figure 10: Mean Improvement in Letters by Viewing Category after accounting for site, age, and prescore

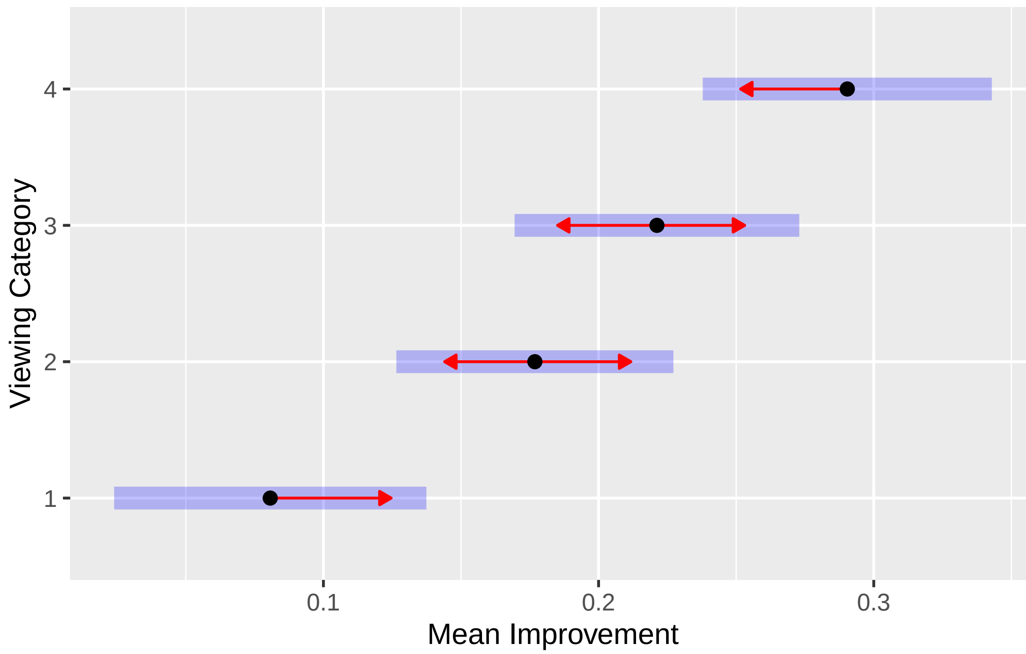


Figure 11: Mean Improvement in Forms by Viewing Category after accounting for site, age, and prescore

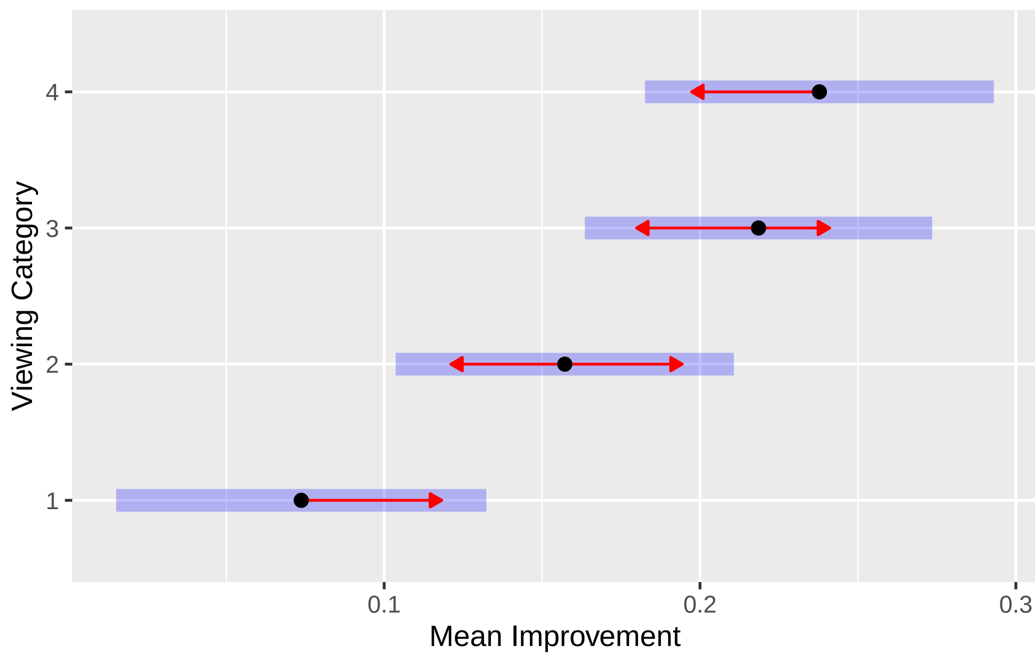


Figure 12: Mean Improvement in Numbers by Viewing Category after accounting for site, age, and prescore

random effect ID). The mathematical form of this model is:

$$\text{Improvement} = \beta_0 + \beta_1 \text{Age} + \beta_{2,3} I_{Subject} + \beta_{4\ldots6} I_{Viewcat} + \beta_{7\ldots10} I_{Site} + \beta_{11\ldots16} I_{Viewcat} I_{Site} + \gamma ID + \epsilon,$$

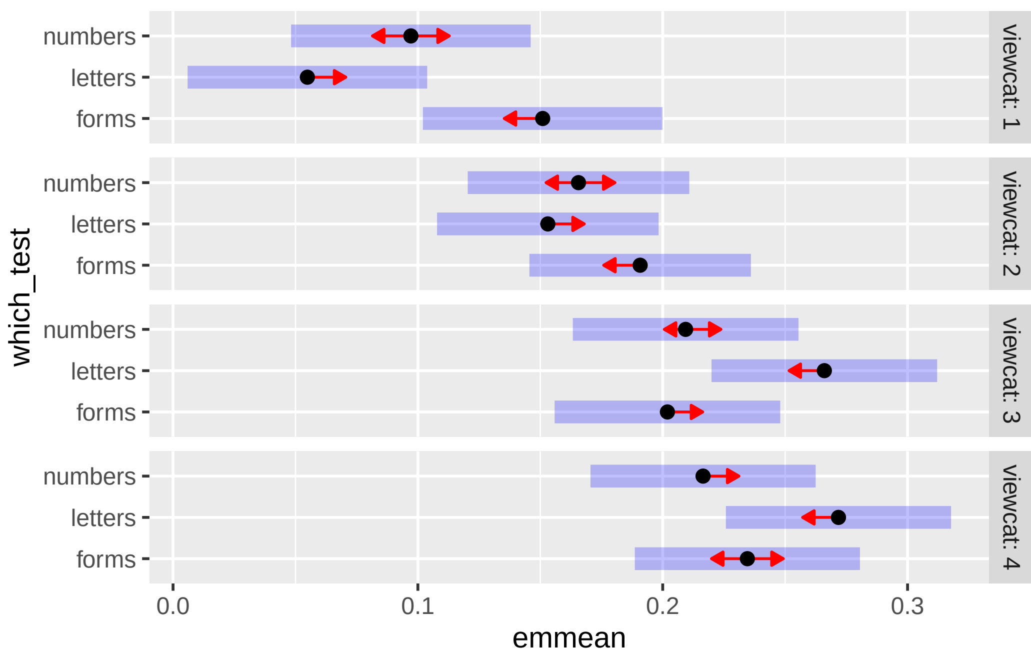
$$\epsilon \sim N(0, \sigma).$$

The interpretations of $\beta 0$ to $\beta 10$ work the same as they did in the previous model. However, in this model we also include an interaction term and a random effects term. The interaction term

```
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: diff_pct
                     Chisq Df Pr(>Chisq)
which_test          12.271  2   0.002165 **
viewcat             33.869  3  2.111e-07 ***
site                14.866  4   0.004987 **
which_test:viewcat 101.107  6  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Linear mixed model fit by REML ['lmerMod']
Formula: diff_pct ~ which_test * viewcat + site + (1 | ID)
   Data: longData

REML criterion at convergence: -2350.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2161 -0.6353 -0.0073  0.6377  3.9825

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 0.01547  0.1244
 Residual             0.01457  0.1207
Number of obs: 2160, groups:  ID, 240

Fixed effects:
                            Estimate Std. Error t value
(Intercept)                 0.118336   0.026291   4.501
which_testletters          -0.096105   0.013413  -7.165
which_testnumbers          -0.053841   0.013413  -4.014
viewcat2                    0.039801   0.027490   1.448
```

```
viewcat3                      0.050941   0.027928   1.824
viewcat4                      0.083597   0.028377   2.946
site2                         0.080962   0.024768   3.269
site3                         0.004154   0.023506   0.177
site4                         0.015741   0.027519   0.572
site5                         0.062351   0.035639   1.749
which_testletters:viewcat2    0.058404   0.018488   3.159
which_testnumbers:viewcat2    0.028656   0.018488   1.550
which_testletters:viewcat3    0.160194   0.018212   8.796
which_testnumbers:viewcat3    0.061277   0.018212   3.365
which_testletters:viewcat4    0.133313   0.018346   7.266
which_testnumbers:viewcat4    0.035741   0.018346   1.948
```

```
Correlation matrix not shown by default, as p = 16 > 12.
Use print(x, correlation=TRUE)  or
    vcov(x)        if you need it
```




```
[1] 6.099354e-190
```