Sesame Street Project Group 1

Jiacheng Chen, Yiyun Gong, Xigang Zhang

10/5/2020

Project Description

The producers of Sesame Street and the Educational Testing Service (ETS) conducted a joint study to evaluate the impact of watching Sesame Street in the United States on children's cognitive performance. The primary objective of Sesame Street is to educate 3-5 years old children in skills related to early school learning, with a special interest in 4-year-old disadvantaged children. ETS conducted an experiment in which children from five different family backgrounds in five different locations in the United States were tested on their knowledge of body parts, letters, forms, numbers, relationship terms, and classification skills. The test was administered in two sessions at least 6 months apart. ETS also measured the pretest of the Peabody Picture Vocabulary for the measure of vocabulary maturity and recorded the age of the children (in months.) and gender, as well as the frequency of viewing, location of viewing, and whether or not they were encouraged to watch Sesame Street.

The client provided part of the raw data from the experiment (see Appendix A for raw data set description) and wanted us to analyze the data to determine if viewing Sesame Street would enhance children's knowledge of body parts, letters, forms, numbers, relational terms, and if there is any area could be focused on for improvement, i.e. are children having less impact on body parts, letters, forms, numbers, relational terms, or classification skills. We selected analyzable variables from the raw data and recorded the difference between the two test results for analysis.

Research Questions need inproment

Question 1: Does watch Sesame Street improve children's cognitive performance on form, number, and letters?

Question 2: If there is any area could be focused on for improvement?

Statistical Questions

To answer client's research questions, we investigated the following statistical questions:

Question 1: What factors have significant impacts on children's test score differences?

Question 2: Which skills of children has the least significant improvement?

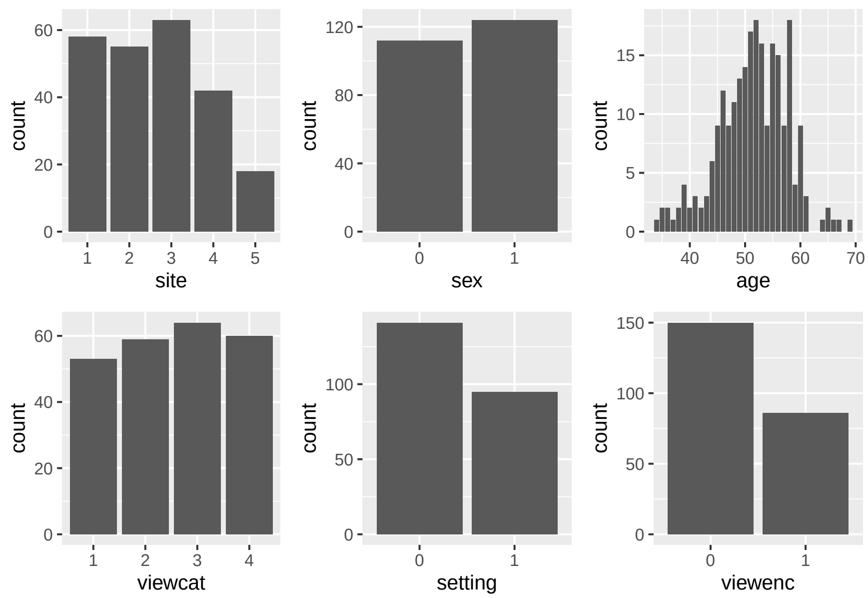
Variables of Interest

We subtracted the pre-test score from the post-test score to obtain the difference between the children's pre and post-Sesame Street test scores. The body, relational terms, and classification skills in variable overlap significantly with forms, numbers, and letters in our analysis (see Appendix A). We will therefore focus on the differences in form, letter, and numbers between children's pre- and post-test scores on watching Sesame Street. The explanatory variables we chose were site, gender, age, viewcat, setting, viewenc, peabody, and body, relational terms, and classification skills differences in pre- and post-test scores.

Variables	Descriptions	Type	Levels
site	Five different sampling sites coded as 1, 2, 3, 4, 5.	Categorical	1 = Three to five-year-old disadvantaged children from inner city areas in various parts of the country. 2 = Four-year-old advantaged suburban children. 3 = Advantaged rural children. 4 = Disadvantaged rural children. 5 = Disadvantaged Spanish speaking children.
sex	Gender of tested children	Categorical	1 = male, 2 = female
age	Tested children's age	Numerical	34 to 69, recorded in months
viewcat	Tested chilrdren's watching frequency.	Categorical	1 = rarely watched the show, to 4 = watched the show on average of more than 5 times a week.
viewenc	Were encouraged to view the show or not.	Categorical	1 = Encouraged, 2 = Not encouraged
setting	Where the tested children watched the show.	Categorical	1 = at home, 2 = in school
peabody	Mental age scores.	Numerical	8-99
diffbody	The differences of test score on body parts.	Numerical	-11 - 16 (original maximum score: 32)
difflet	The differences of test score on letters.	Numerical	-22 - 35 (original maximum score: 58)
diffform	The differences of test score on forms.	Numerical	-10 - 15 (original maximum score: 20)
diffnumb	The differences of test score on numbers.	Numerical	-35 - 32 (original maximum score: 54)
diffrel	The differences of test score on relational terms.	Numerical	-10 - 7 (original maximum score: 17)
diffclasf	The differences of test score on classification skills.	Numerical	-7 - 12 (original maximum score: 24)

Exploratory Data Analysis (EDA)

Our data were reviewed and pre-processed prior to the statistical analysis (See Appendix B for raw dataset analysis and pre-processing producers). From raw dataset exploration, we noticed that a total of 4 different records on 3 different post-exams had exceeded the maximum score. Thus, we decided to omit these 4 records to maintain our model accuracy. Therefore, our sample size reduces from 240 to 236. Our main goal on this research is to focus on whether watching Sesame Street has a significant influence on children's performance. Therefore, we calculated the score difference between post-tests and pre-tests for each exam which reduces our columns on tests from 12 columns (6 pre-test+6 post-test) to 6 score difference columns. Besides, there were no notable data errors or missing values. We want to Therefore, First thing first, we would like to know the distribution of all categorical variables, see Figure1.



`Figure 1`: Histogram on Variables except Test scores. Notably, while site5 have the least sample, sex, viewcat, is equal and age is normally distributed.

In `Figure 1` while site5 have the least sample, sex, viewcat, is equal and age is normally distributed. Then we looked at the descriptive summary statistics, shown in `Table 2`, on each categorical variable for sites which are the five populations were of interest that our client initially adopted and interested in.

Characteristic	**N**	**1**, N = 58	**2**, N = 55	**3**, N = 63	**4**, N = 42	**5**, N = 18
sex, n (%)	236					
0		30 (52%)	24 (44%)	30 (48%)	18 (43%)	10 (56%)
1		28 (48%)	31 (56%)	33 (52%)	24 (57%)	8 (44%)
agecat, n (%)	236					
0		33 (57%)	26 (47%)	26 (41%)	19 (45%)	9 (50%)
1		25 (43%)	29 (53%)	37 (59%)	23 (55%)	9 (50%)
viewcat, n (%)	236					
1		9 (16%)	4 (7.3%)	12 (19%)	22 (52%)	6 (33%)
2		16 (28%)	10 (18%)	17 (27%)	10 (24%)	6 (33%)
3		20 (34%)	17 (31%)	20 (32%)	6 (14%)	1 (5.6%)
4		13 (22%)	24 (44%)	14 (22%)	4 (9.5%)	5 (28%)
setting, n (%)	236					
0		28 (48%)	31 (56%)	46 (73%)	18 (43%)	18 (100%)
1		30 (52%)	24 (44%)	17 (27%)	24 (57%)	0 (0%)
viewenc, n (%)	236					

Characteristic	**N**	**1**, N = 58	**2**, N = 55	**3**, N = 63	**4**, N = 42	**5**, N = 18
0		31 (53%)	36 (65%)	49 (78%)	20 (48%)	14 (78%)
1		27 (47%)	19 (35%)	14 (22%)	22 (52%)	4 (22%)

`Table 2`: The table displays each categorical variable with the count and percentage and group by each population of interest. There were a total of 236 valid observations in this data. No records were incomplete. Notably, there are 0 records for setting 1 (viewed at school) on site 5 (Disadvantaged Spanish Speaking children). Furthermore, because site 5 has a much smaller amount of sample records than site 1-4, the amount of records for each categorical variable for site 5 is relatively smaller.

Next, we looked at the descriptive summary statistics on score difference between post-test and pre-test for tests on body parts, letters, forms, numbers, relations, and classification for each site (site1-5), shown in `Table 3`.

Test Score Differences on Each Population of Interest

	1 (N=58)	**2** (N=55)	**3** (N=63)	**4** (N=42)	**5** (N=18)	**Overall** (N=236)
body parts						
Mean (SD)	2.95 (6.00)	2.82 (4.01)	4.43 (5.14)	4.83 (5.09)	5.61 (3.35)	3.85 (5.07)
Median [Min, Max]	2.00 [-11.0, 16.0]	3.00 [-8.00, 10.0]	5.00 [-9.00, 15.0]	5.00 [-4.00, 23.0]	6.00 [0, 10.0]	4.00 [-11.0, 23.0]
letters						
Mean (SD)	10.3 (12.3)	19.6 (10.4)	6.63 (8.31)	7.02 (9.41)	9.67 (8.84)	10.8 (11.2)
Median [Min, Max]	7.00 [-22.0, 35.0]	19.0 [-1.00, 41.0]	5.00 [-17.0, 28.0]	6.50 [-7.00, 30.0]	8.00 [-3.00, 26.0]	9.00 [-22.0, 41.0]
forms						
Mean (SD)	3.21 (4.01)	4.31 (3.10)	4.19 (4.32)	3.12 (3.54)	4.89 (2.45)	3.84 (3.74)
Median [Min, Max]	3.00 [-10.0, 15.0]	4.00 [-2.00, 12.0]	4.00 [-10.0, 17.0]	3.00 [-5.00, 11.0]	5.00 [0, 9.00]	4.00 [-10.0, 17.0]
numbers						
Mean (SD)	7.79 (12.4)	12.5 (8.50)	8.75 (8.87)	6.67 (7.86)	10.1 (8.61)	9.12 (9.75)
Median [Min, Max]	7.00 [-35.0, 32.0]	13.0 [-4.00, 28.0]	9.00 [-12.0, 33.0]	6.50 [-7.00, 27.0]	6.50 [-1.00, 33.0]	9.00 [-35.0, 33.0]
relations						
Mean (SD)	1.09 (3.70)	1.44 (2.59)	2.41 (3.46)	2.17 (3.97)	0.556 (1.85)	1.67 (3.38)
Median [Min, Max]	2.00 [-10.0, 7.00]	2.00 [-7.00, 10.0]	3.00 [-7.00, 12.0]	1.50 [-5.00, 13.0]	0.500 [-4.00, 4.00]	2.00 [-10.0, 13.0]
classification						
Mean (SD)	3.34 (4.14)	4.53 (4.09)	3.35 (4.86)	2.60 (4.56)	4.28 (4.81)	3.56 (4.47)
Median [Min, Max]	3.00 [-7.00, 12.0]	4.00 [-5.00, 12.0]	4.00 [-7.00, 14.0]	2.00 [-6.00, 14.0]	7.00 [-4.00, 11.0]	3.50 [-7.00, 14.0]

`Table 3`: The table displays each of the score differences calculated from the subtraction between post-test scores and pre-tests scores. There were no incomplete records for any tests. Thus, we have the same sample sizes for each test overall. The table provides statistical summaries on mean, standard deviation, median, min, max for each test, and overall summary statistics on each population (site).

Notably, from `Table 3`, most of the test difference ranges from a negative value to positive value, except for body parts test and forms test on site 5 (Disadvantaged Spanish Speaking Children). However, the cause of this situation might be varied. (See Limitation Section for additional explanation).

Statistical Analysis

Analytical Methods

To answer the first research question, whether viewing Sesame Street improves children's knowledge of `letters`, `numbers`, and `forms`, we employed **Two-Way Anova Test** and **Multiple Linear Regression Models**, three times for each aspect. We factored our categorical variables: `site, sex, viewcat, setting` to see if certain type of variable will have significant difference. We then analyzed rest numerical variables to checked assumption for Multi-linear Regression Model(see Appendix C)

We use stepwise method to select the best fitted model for three cognitive skills. The stepwise method is fitting regression models in which the choice of predictive variables is carried out by an automatic procedure. In each step, the method selected the model with the lowest AIC (Akaike Information Criterion) score and a variable is considered for addition to or subtraction from the set of explanatory variables based on some pre-specified criterion.

The AIC score is a criterion for evaluating model quality. When given a set of models with data, the AIC estimates the quality of each model, relative to every other model. The lower the AIC score, the better the model fits the experimental data.

Final Models

The first Multi-linear model was built to show the impact of selected variables on the improvement of number after watching sesame st. Using stepwise method and multivariate test we found the final model included `prenumb, diffclasf, diffrelat, site, age, viewcat`, interaction between `age:site, site:prelet, age:peabody, prelet:peabody`. As the client was interested in if `encouragement` will influence the outcome, we also add it into the model. (See Appendix C for final model). The regression coefficients produced for the difference of knowledge of numbers was found to be:

Variables	Coefficients
(Intercept)	-6.9184789
age	0.2015555
viewcat2	2.9359733
viewcat3	5.3054618
viewcat4	5.1084374
site2	4.2229346
site3	0.9574310
site4	0.7140068
site5	3.9916552
prenumb	-0.4428552
preform	0.8224894
diffclasf	1.0035243
encour	-1.3493048
age:peabody	-0.0003514
peabody:prelet	-0.0003672

Table 4: Regression Coefficients for the difference of knowledge of **numbers**.

Same idea to letter and form, while the final model for letter include, `viewcat`, `prenumb`, interaction between `prenumb:prelet`,`prelet:diffbody`,`prelet:diffclasf`,`diffclasf:preform`. While client also want to know if `site, setting, sex, age` have correlation with outcome, we add those variables to final model.

Variables	Coefficients
(Intercept)	-7.266136
site2	6.216937
site3	-5.115192
site4	-0.989296
site5	0.182271
viewcat2	3.525912
viewcat3	9.514882
viewcat4	8.866553
age	0.203952
encour	1.597915
prenumb:prelet	-0.004641
prelet:diffbody	0.017539
prelet:diffclasf	-0.039543
diffclasf:preform	0.120070

Table 5: Regression Coefficients for the difference of knowledge of **letters**.

For the final model of form, variables include, `age, viewcat, preform, improverelat`, still the same reason as previous, we add `sex, site, setting, age, encour` to the final model for client interests.

Variables	Coefficients
(Intercept)	1.79004
viewcat2	1.46667
viewcat3	2.26445
viewcat4	3.25229
age	0.12087
site2	1.21784
site3	-1.42899
site4	-0.56138
site5	0.77210
preform	-0.64405
diffrel	0.28820
encour	0.07038

Table 6: Regression Coefficients for the difference of knowledge of **forms**.

The second question we want to look at is if there is an area that could be focused on for improvement. Thus, we looked at the results from `Multiple linear regression` and `ANOVA` to experiment the significant level of each explanatory variable. For the first model for number test, we noticed that the interaction between `peabody` and `prelet` was the least significant variable. For the second model for letter improvement, `age` was the least significant variable in this model. For the third model for form improvement, if or not being `encouraged` to watch the show is the least significant variable.

Recommendations

The main research questions our client wanted to answer was "Does watch Sesame Street improve children's cognitive performance on form, number, and letters?"

Through our analysis, each of the view categories had a positive impact on the differences in children's performance in numbers, letters, and forms. The p-values were all less than 0.05, indicating that these effects were significantly meaningful. Therefore, we conclude that viewing Sesame Street could improve preschool children's cognitive abilities in numbers, letters, and forms.

The client's second question was "If there is any area could be focused on for improvement?"

Through the results from `Multiple linear regression` and `ANOVA` to experiment the significant level of each explanatory variable. we consider that The factors that need attention for different cognitive skills are different.

We found that for children with significantly improved classification ability, there was also a significant improvement in knowledge of numbers after watching Sesame Street. We also noted that peabody scores and initial knowledge of letters had little effect on children's ability to improve their knowledge of numbers. We therefore suggest that the production team should focus the content on classification in order to improve children's knowledge of numbers.

We found that children who have improvement on body parts, and classification skills will improve a lot in their knowledge of letters. The least significant factor in improving the cognitive ability of letters was age. We therefore recommend that the production team add more content about common objects and their classification in the program to improve children's knowledge of letters more effectively.

As for how to improve children's knowledge of forms, we found that children who improved significantly in knowledge of relational terms also improved significantly in knowledge of forms. The effect of whether children were encouraged to watch the program was not significant. Therefore, we suggest that the production team consider focusing the content of the program on teaching the knowledge of relational terms such as amount, size, and position.

Additional Considerations

During the project, we noticed there are some limitations needed further attention:

First, from our categorical summaries table (See Table 2), The sample size of Site 5 (Disadvantaged Spanish Speaking children) is relatively small (n=18) compared to other sites. It might not be considered as a large sample size to follow the normal distribution. Therefore, it may cause inaccurate conclusions on our study on site 5.

Second, there were four different records on three different post-exams that had exceeded the maximum score. The reasons to cause this error might be various. Without further information, we cannot make certain assumptions on the cause of the error.

Third, we also face the possibility that the samples were not randomly selected, and some sampling errors may be involved in the experiment design. We did not have further information from our client to check if data was collected randomly and appropriately. We cannot rule out the possibility that sampling error might be involved in our data.

We would like to thank the Sesame Street producing team and the Educational Testing Service (ETS)for the opportunity to work on this project. We enjoyed it and hope our results could be helpful to improve the Sesame Street. We appreciate our classmates who provided many wonderful suggestions for helping us improve this report.

R-Package:

car:

Fox J, Weisberg S (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

stargazer:

Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. https://CRAN.R-project.org/package=stargazer

tidyverse:

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019).

"Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.

MASS:

Venables WN, Ripley BD (2002). Modern Applied Statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0, http://www.stats.ox.ac.uk/pub/MASS4/.

corrplot:

Taiyun Wei and Viliam Simko (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot

gtsummary:

Daniel D. Sjoberg, Margie Hannum, Karissa Whiting and Emily C. Zabor (2020). gtsummary: Presentation-Ready Data Summary and Analytic Result Tables. R package version 1.3.4. https://CRAN.R-project.org/package=gtsummary

Table1:

Benjamin Rich (2020). table1: Tables of Descriptive Statistics in HTML. R Table1:package version 1.2. https://CRAN.R-project.org/package=table1

Document Reference:

Penn State University. (n.d.). 10.2.1 - ANOVA Assumptions | STAT 500. PennState: Statistics Online Courses. Retrieved September 21, 2020, from https://online.stat.psu.edu/stat500/lesson/10/10.2/10.2.1

StatisticsSolutions. (n.d.). Assumptions of Multiple Linear Regression. Retrieved September 21, 2020, from https://www.statisticssolutions.com/assumptions-of-multiple-linear-regression/

Tutorial: tbl_summary. (2020). Retrieved 21 September 2020, from http://www.danieldsjoberg.com/gtsummary/articles/tbl_summary.html

table1 package v1.2, t. (2020). | R Documentation. Retrieved 21 September 2020, from https://www.rdocumentation.org/packages/table1/versions/1.2

Technical Appendix

Appendix A: Raw Data Description

The original data set contains 240 objects and 23 variables. Besides the variables we used in our analysis, there are some variables that we did not use in our analysis, and some variables that we have modified to better build the model.

Variable Name	Description	Reason to remove
agecat	Whether the age is greater than 51 month.	Not significant in model.
encour	Whether the child is encouraged to watch Sesame Street.	Overlapped with viewenc.
regular	We are not sure what this variable represents.	Not significant in model.
pre-scores	Children's performance on tests prior to watching Sesame Street.	We decided to use the difference
post-scores	Children's performance on tests after watching Sesame Street.	between the two tests in the final model.

In Exploratory Data Analysis, we found that some objects did not have the same scale of pre- and post-test scores. We analyzed them and found that several objects had post-test scores for body parts, letters, and relational terms that exceeded the maximum value of the scale. Therefore, we decided to remove these objects from the modeling to maintain the accuracy of the post-test analysis.

Appendix B Additional Tables and Figures used for EDA

Characteristic	**N**	**1**, N = 60	**2**, N = 55	**3**, N = 64	**4**, N = 43	**5**, N = 18
sex, n (%)	240					
1		31 (52%)	24 (44%)	31 (48%)	19 (44%)	10 (56%)
2		29 (48%)	31 (56%)	33 (52%)	24 (56%)	8 (44%)
agecat, n (%)	240					
1		34 (57%)	26 (47%)	27 (42%)	19 (44%)	9 (50%)
2		26 (43%)	29 (53%)	37 (58%)	24 (56%)	9 (50%)
viewcat, n (%)	240					
1		9 (15%)	4 (7.3%)	12 (19%)	23 (53%)	6 (33%)
2		17 (28%)	10 (18%)	17 (27%)	10 (23%)	6 (33%)
3		20 (33%)	17 (31%)	20 (31%)	6 (14%)	1 (5.6%)
4		14 (23%)	24 (44%)	15 (23%)	4 (9.3%)	5 (28%)
setting, n (%)	240					
1		28 (47%)	31 (56%)	47 (73%)	19 (44%)	18 (100%)
2		32 (53%)	24 (44%)	17 (27%)	24 (56%)	0 (0%)
viewenc, n (%)	240					
1		32 (53%)	36 (65%)	50 (78%)	20 (47%)	14 (78%)
2		28 (47%)	19 (35%)	14 (22%)	23 (53%)	4 (22%)

Table B1: The table displays the summaries on the number and percentage of each categorical variables on each population (sites) in our original dataset.

Raw Data Summaries on Each Pre-test and Post-test

	1 (N=60)	2 (N=55)	3 (N=64)	4 (N=43)	5 (N=18)	Overall (N=240)
prebody						
Mean (SD)	22.0 (6.26)	26.1 (4.46)	18.1 (6.21)	20.0 (5.98)	20.6 (4.87)	21.4 (6.39)
Median [Min, Max]	23.0 [11.0, 32.0]	27.0 [17.0, 32.0]	17.0 [6.00, 31.0]	20.0 [8.00, 31.0]	20.0 [15.0, 30.0]	22.0 [6.00, 32.0]
postbody						
Mean (SD)	24.9 (5.02)	28.9 (2.99)	22.6 (6.47)	24.7 (5.06)	26.2 (4.49)	25.3 (5.50)
Median [Min, Max]	25.5 [12.0, 32.0]	29.0 [18.0, 32.0]	23.0 [11.0, 39.0]	25.0 [13.0, 32.0]	28.0 [15.0, 31.0]	27.0 [11.0, 39.0]
prelet						
Mean (SD)	17.7 (11.6)	17.8 (8.09)	13.3 (4.40)	14.4 (5.67)	17.6 (12.0)	15.9 (8.54)
Median [Min, Max]	14.0 [2.00, 48.0]	16.0 [6.00, 46.0]	13.0 [4.00, 28.0]	14.0 [1.00, 30.0]	15.5 [5.00, 55.0]	14.0 [1.00, 55.0]
postlet						
Mean (SD)	28.0 (14.6)	37.4 (10.7)	20.0 (9.16)	21.2 (10.8)	27.2 (13.8)	26.7 (13.4)
Median [Min, Max]	23.0 [6.00, 63.0]	40.0 [15.0, 53.0]	16.5 [6.00, 46.0]	17.0 [0, 46.0]	23.0 [11.0, 54.0]	23.0 [0, 63.0]
preform						
Mean (SD)	10.8 (3.99)	12.0 (3.04)	7.95 (2.92)	9.40 (3.58)	8.89 (3.85)	9.93 (3.73)
Median [Min, Max]	10.0 [2.00, 19.0]	12.0 [4.00, 18.0]	8.00 [2.00, 14.0]	9.00 [2.00, 17.0]	8.00 [2.00, 19.0]	10.0 [2.00, 19.0]
postform						
Mean (SD)	14.1 (3.63)	16.3 (2.83)	12.1 (4.41)	12.4 (3.74)	13.8 (3.32)	13.7 (4.00)
Median [Min, Max]	14.0 [6.00, 20.0]	17.0 [9.00, 20.0]	12.5 [3.00, 20.0]	13.0 [0, 19.0]	12.5 [9.00, 19.0]	14.0 [0, 20.0]
prenumb						
Mean (SD)	22.2 (13.2)	26.0 (9.50)	16.6 (7.19)	18.5 (8.81)	21.3 (12.2)	20.9 (10.7)
Median [Min, Max]	19.0 [4.00, 52.0]	24.0 [11.0, 48.0]	16.0 [4.00, 37.0]	18.0 [1.00, 44.0]	18.5 [4.00, 52.0]	19.0 [1.00, 52.0]
postnumb						
Mean (SD)	30.1 (13.9)	38.5 (10.7)	25.4 (10.4)	25.2 (10.5)	31.4 (15.0)	30.0 (12.8)
Median [Min, Max]	28.0 [6.00, 54.0]	41.0 [8.00, 53.0]	23.5 [0, 46.0]	25.0 [0, 44.0]	34.5 [10.0, 54.0]	29.0 [0, 54.0]
prerelat						
Mean (SD)	10.4 (3.57)	11.5 (2.20)	8.56 (2.78)	9.07 (2.93)	10.7 (2.30)	9.94 (3.07)
Median [Min, Max]	10.0 [2.00, 17.0]	11.0 [5.00, 15.0]	9.00 [3.00, 15.0]	9.00 [3.00, 14.0]	10.0 [8.00, 15.0]	10.0 [2.00, 17.0]
postrelat						
Mean (SD)	11.6 (3.00)	12.9 (2.38)	11.0 (2.70)	11.5 (3.86)	11.2 (2.51)	11.7 (3.00)
Median [Min, Max]	11.0 [4.00, 19.0]	13.0 [7.00, 17.0]	11.0 [1.00, 16.0]	11.0 [0, 23.0]	11.5 [6.00, 15.0]	12.0 [0, 23.0]
preclasf						
Mean (SD)	13.0 (5.62)	14.7 (3.94)	10.3 (3.49)	11.1 (4.04)	11.4 (4.29)	12.2 (4.64)
Median [Min, Max]	13.0 [1.00, 24.0]	15.0 [7.00, 24.0]	10.0 [4.00, 17.0]	11.0 [0, 20.0]	11.0 [4.00, 23.0]	12.0 [0, 24.0]
postclasf						
Mean (SD)	16.2 (4.94)	19.3 (3.45)	13.7 (4.92)	13.5 (4.78)	15.7 (5.94)	15.7 (5.15)
Median [Min, Max]	16.5 [6.00, 24.0]	19.0 [7.00, 24.0]	15.0 [4.00, 22.0]	13.0 [0, 23.0]	17.5 [7.00, 23.0]	16.0 [0, 24.0]

Table B2: The table displays summaries statistics on raw data on each pre-test and post-test for all 5 sites (population) with information on Mean, Standard deviation, median, min, and max. We can notice that some of the maximum scores on post-tests are exceeding the max score described on the data description portion.

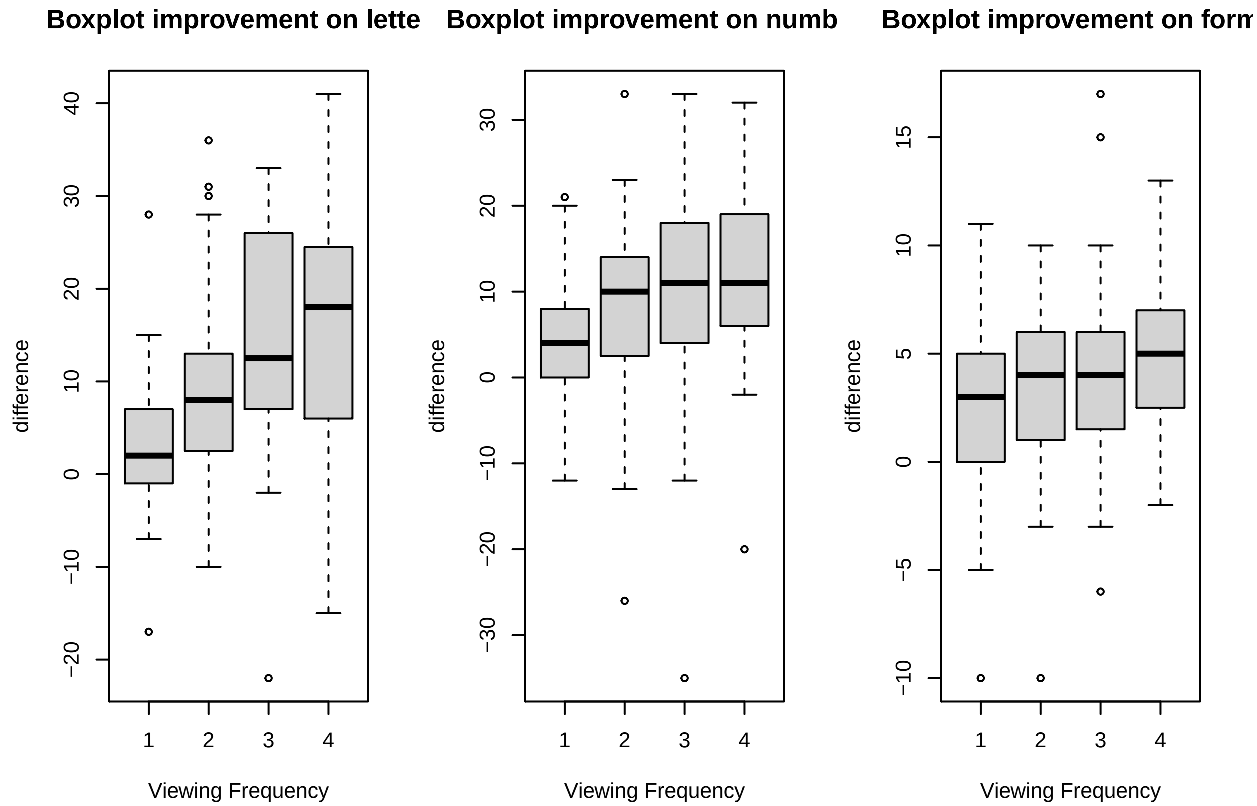


Figure B1: Boxplot of score improvement with respect to letter, number, form, while explanatory variable is viewcat(view frequency). It reflects there are positive relationships between viewcat with improved scores especially in letter aspect.

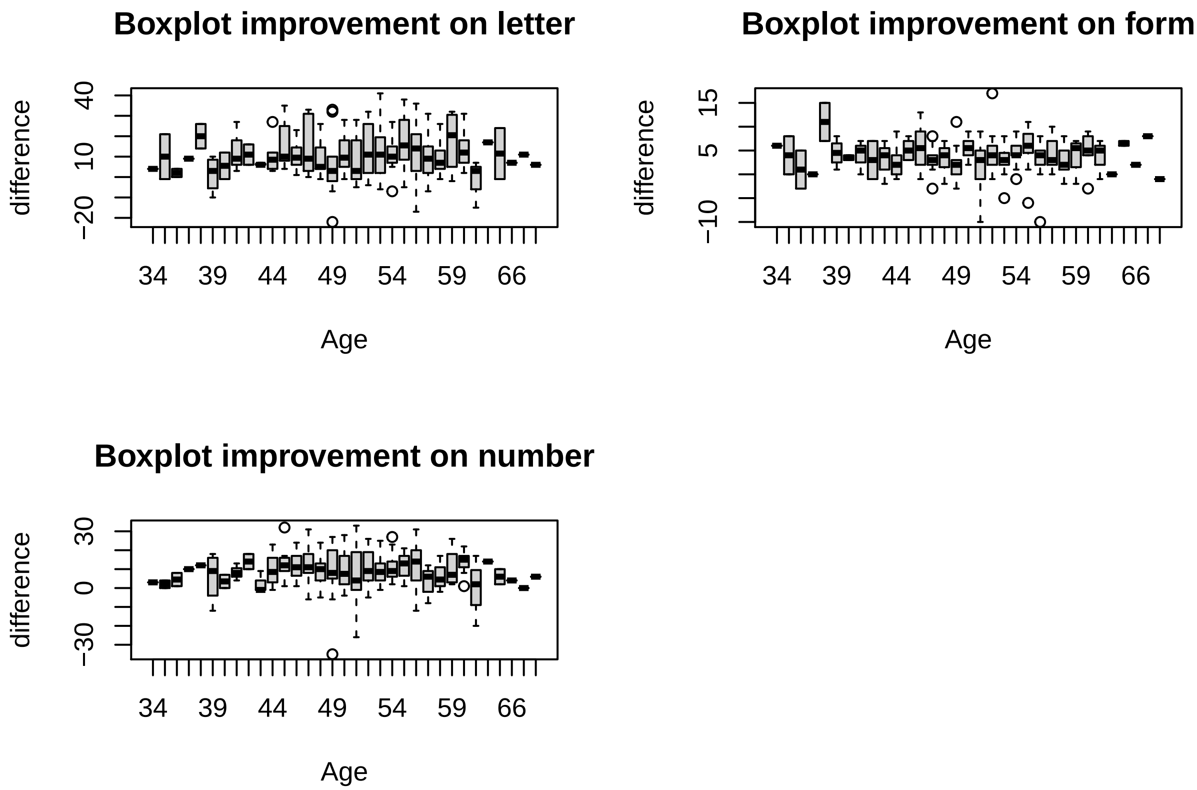


Figure B2: Boxplot of score improvement with respect to age with scores. We notice that there is no obvious trend indicating that the older children will have more improvement in scores."

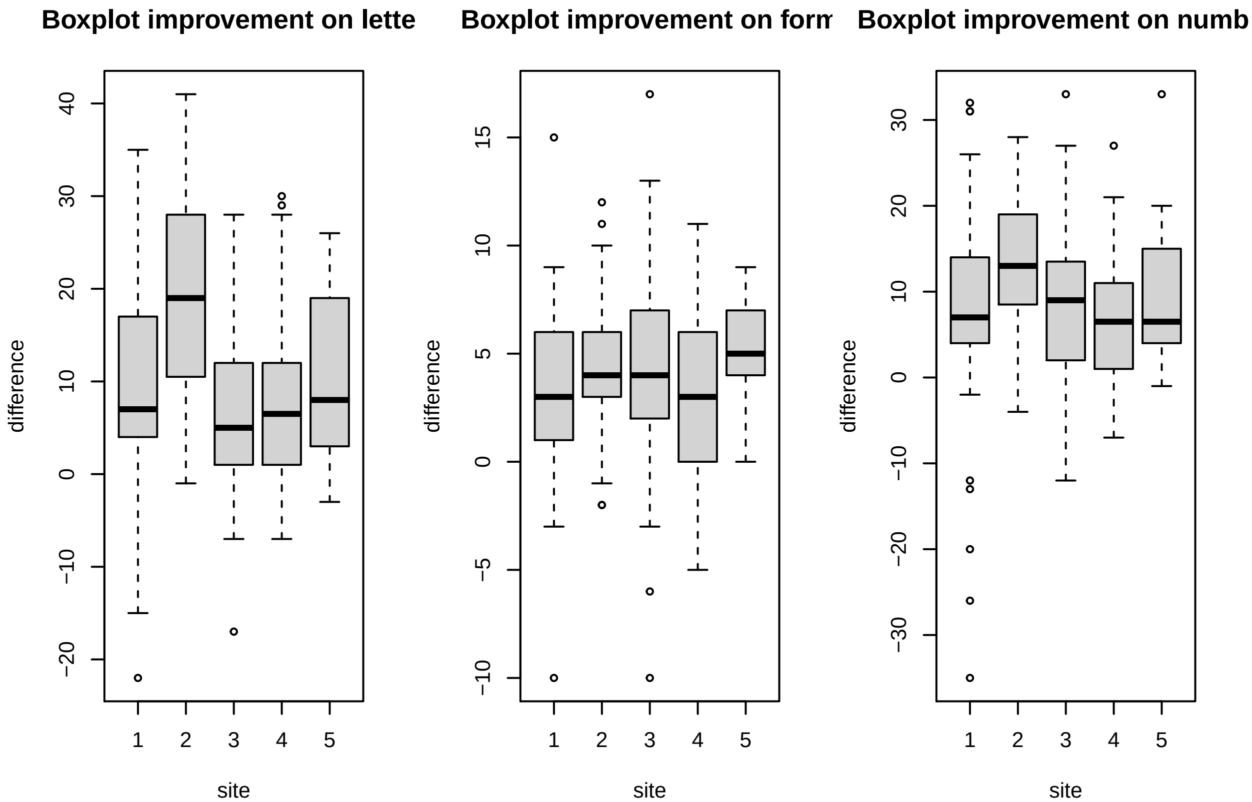


Figure B3: Boxplot of score improvement with respect to whether site difference is a confounder that influences scores difference. It reflects that except site2, there is no significant difference indicating sites have strong correlation with improvement in scores.

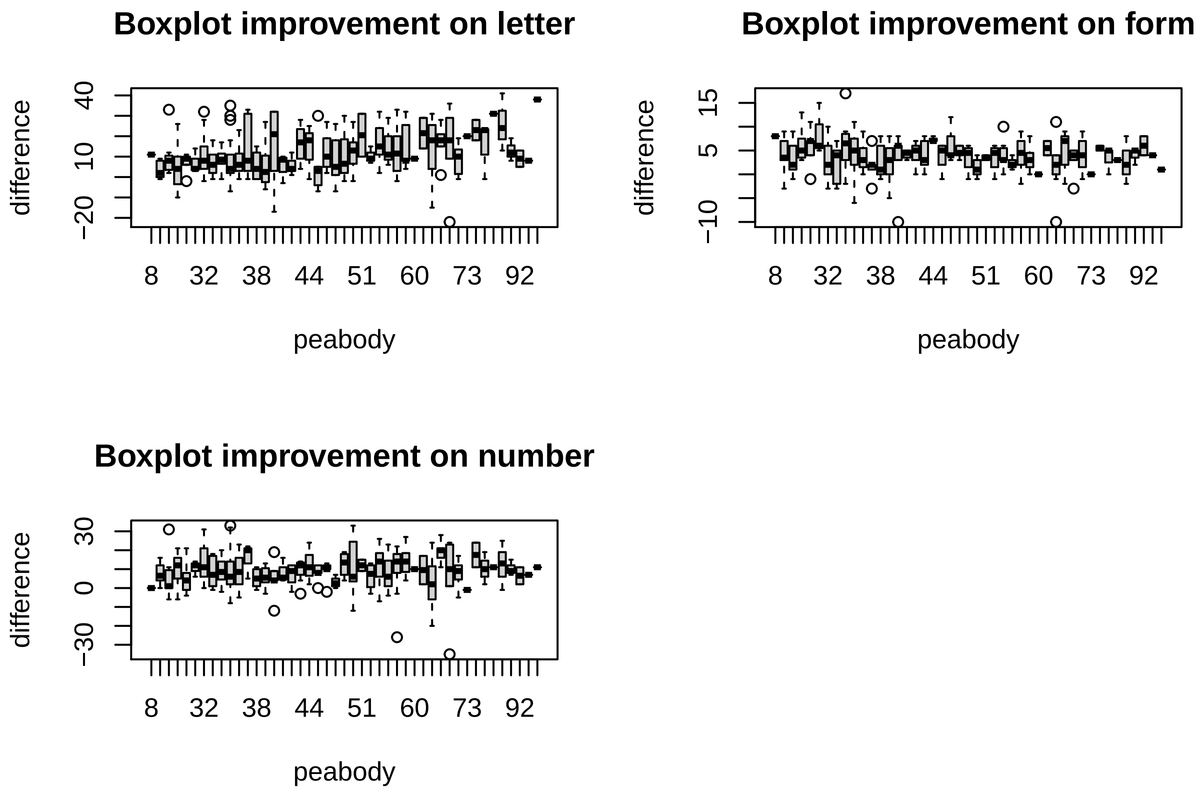


Figure B4: Boxplot of score improvement with respect to if peabody as a confounder to influence improvement in scores. Which we noticed that there is no obvious trend to say higher peabody scores will correlate with higher improvement in scores.

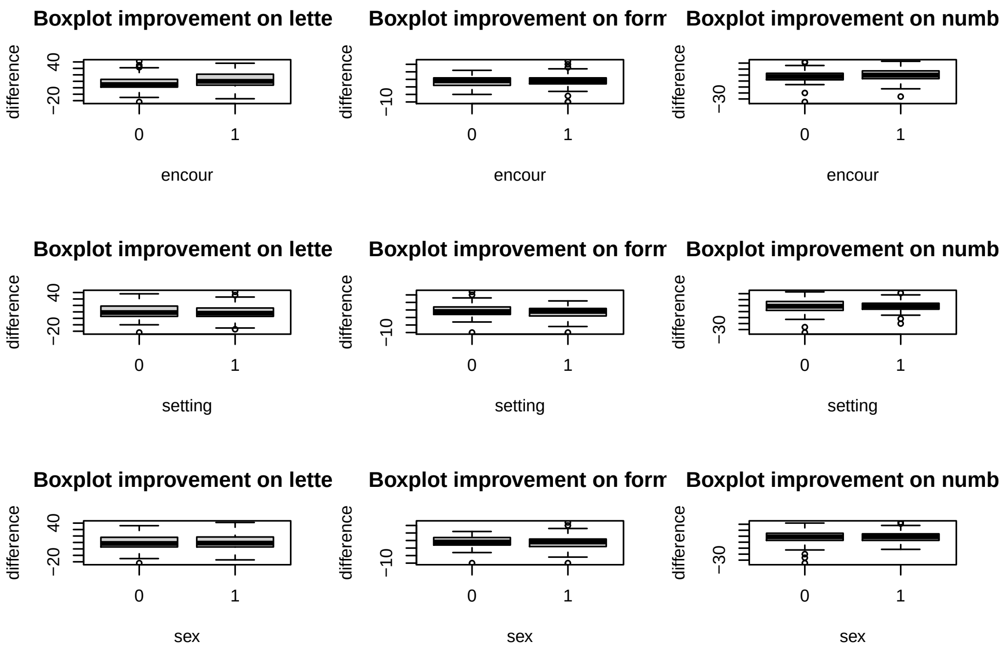


Figure B5: Boxplot of score improvement with respect to sex. Indicating there is no difference between sex and improvement in scores.

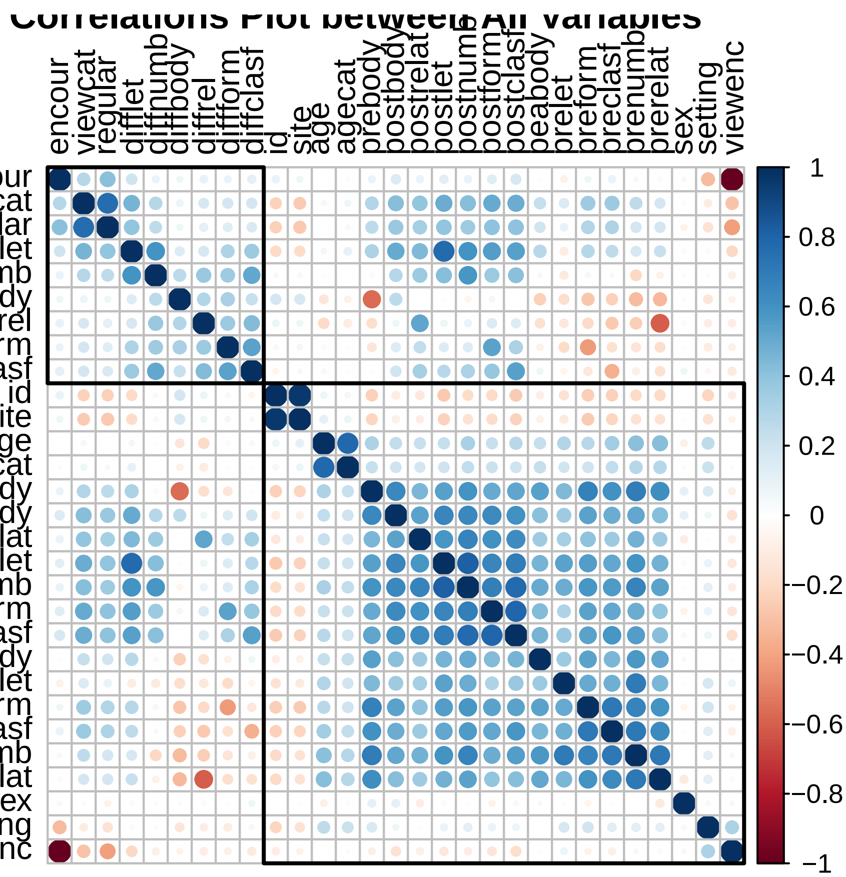


Figure B6: Correlation plot with all variables.

Appendix C Assumptions Checked for Each of the Models

Final Models for Number

```
##
## Call:
## lm(formula = diffnumb ~ age + viewcat + site + prenumb + preform +
##     diffclasf + age:peabody + prelet:peabody + encour, data = sesame2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.6558  -4.8485   0.0148   5.1088  18.1192
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6.9184789  4.6036016  -1.503  0.13431
## age             0.2015555  0.1023668   1.969  0.05021 .
## viewcat2        2.9359733  1.6206015   1.812  0.07140 .
## viewcat3        5.3054618  1.6649250   3.187  0.00165 **
## viewcat4        5.1084374  1.7367778   2.941  0.00362 **
## site2           4.2229346  1.6017307   2.636  0.00897 **
## site3           0.9574310  1.5180664   0.631  0.52889
## site4           0.7140068  1.6300238   0.438  0.66179
## site5           3.9916552  2.1901339   1.823  0.06972 .
## prenumb        -0.4428552  0.0830819  -5.330 2.41e-07 ***
## preform         0.8224894  0.2007006   4.098 5.85e-05 ***
## diffclasf       1.0035243  0.1171860   8.564 1.88e-15 ***
## encour         -1.3493048  1.1823359  -1.141  0.25501
## age:peabody    -0.0003514  0.0009193  -0.382  0.70265
## peabody:prelet -0.0003672  0.0013880  -0.265  0.79159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.526 on 221 degrees of freedom
## Multiple R-squared:  0.4391, Adjusted R-squared:  0.4036
## F-statistic: 12.36 on 14 and 221 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: diffnumb
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## age              1     4.6     4.6  0.0810   0.77622
## viewcat          3  1897.9   632.6 11.1684 7.473e-07 ***
## site             4   492.8   123.2  2.1748   0.07278 .
## prenumb          1  2858.9  2858.9 50.4692 1.643e-11 ***
## preform          1   367.2   367.2  6.4815   0.01158 *
## diffclasf        1  4094.2  4094.2 72.2784 2.809e-15 ***
## encour           1    63.2    63.2  1.1166   0.29181
## age:peabody      1    17.3    17.3  0.3049   0.58139
## peabody:prelet   1     4.0     4.0  0.0700   0.79159
## Residuals      221 12518.7    56.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final Models for Letter

```
##
## Call:
## lm(formula = difflet ~ site + viewcat + age + prenumb:prelet +
##     prelet:diffbody + prelet:diffclasf + diffclasf:preform +
##     encour, data = sesame2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.1577  -5.0223  -0.2988   4.2081  21.3676
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -7.266136   4.749763  -1.530 0.127493
## site2              6.216937   1.586348   3.919 0.000118 ***
## site3             -5.115192   1.607896  -3.181 0.001676 **
## site4             -0.989296   1.792339  -0.552 0.581532
## site5              0.182271   2.342942   0.078 0.938060
## viewcat2           3.525912   1.743215   2.023 0.044309 *
## viewcat3           9.514882   1.776582   5.356 2.12e-07 ***
## viewcat4           8.866553   1.841139   4.816 2.71e-06 ***
## age                0.203952   0.093053   2.192 0.029434 *
## encour             1.597915   1.270242   1.258 0.209727
## prenumb:prelet    -0.004641   0.001492  -3.111 0.002107 **
## prelet:diffbody    0.017539   0.007138   2.457 0.014765 *
## prelet:diffclasf  -0.039543   0.015604  -2.534 0.011959 *
## diffclasf:preform  0.120070   0.026909   4.462 1.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.191 on 222 degrees of freedom
## Multiple R-squared:  0.4977, Adjusted R-squared:  0.4683
## F-statistic: 16.92 on 13 and 222 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: difflet
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## site                4  5973.1 1493.28 22.2589 1.815e-15 ***
## viewcat             3  4464.8 1488.26 22.1841 1.332e-12 ***
## age                 1    76.8   76.81  1.1450  0.285766
## encour              1   246.5  246.49  3.6743  0.056543 .
## prenumb:prelet      1  1431.2 1431.21 21.3338 6.536e-06 ***
## prelet:diffbody     1   604.8  604.75  9.0144  0.002985 **
## prelet:diffclasf    1   625.1  625.07  9.3173  0.002547 **
## diffclasf:preform   1  1335.7 1335.68 19.9097 1.291e-05 ***
## Residuals         222 14893.3   67.09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final Models for Form

```
##
## Call:
## lm(formula = diffform ~ viewcat + age + site + preform + diffrel +
##     encour, data = sesame2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4796 -2.0061  0.3739  1.8737  7.6456
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.79004    1.67308   1.070 0.285811
## viewcat2     1.46667    0.61928   2.368 0.018719 *
## viewcat3     2.26445    0.63803   3.549 0.000471 ***
## viewcat4     3.25229    0.67557   4.814 2.72e-06 ***
## age          0.12087    0.03306   3.656 0.000319 ***
## site2        1.21784    0.55592   2.191 0.029507 *
## site3       -1.42899    0.58067  -2.461 0.014614 *
## site4       -0.56138    0.62963  -0.892 0.373564
## site5        0.77210    0.81737   0.945 0.345874
## preform     -0.64405    0.06360 -10.127  < 2e-16 ***
## diffrel      0.28820    0.06104   4.722 4.12e-06 ***
## encour       0.07038    0.44546   0.158 0.874604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.89 on 224 degrees of freedom
## Multiple R-squared:  0.4321, Adjusted R-squared:  0.4043
## F-statistic:  15.5 on 11 and 224 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: diffform
##            Df  Sum Sq Mean Sq  F value    Pr(>F)
## viewcat     3  117.74   39.25   4.6999  0.003337 **
## age         1    4.20    4.20   0.5033  0.478808
## site        4   65.74   16.44   1.9682  0.100299
## preform     1 1048.48 1048.48 125.5619 < 2.2e-16 ***
## diffrel     1  187.04  187.04  22.3993 3.924e-06 ***
## encour      1    0.21    0.21   0.0250  0.874604
## Residuals 224 1870.47    8.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                    GVIF Df GVIF^(1/(2*Df))
## age             1.681585  1        1.296759
## viewcat         1.769384  3        1.099775
## site            2.256303  4        1.107069
## prenumb         3.212994  1        1.792483
## preform         2.338585  1        1.529243
## diffclasf       1.138977  1        1.067228
## encour          1.348943  1        1.161440
## age:peabody     3.123589  1        1.767368
## peabody:prelet  3.153305  1        1.775755
```

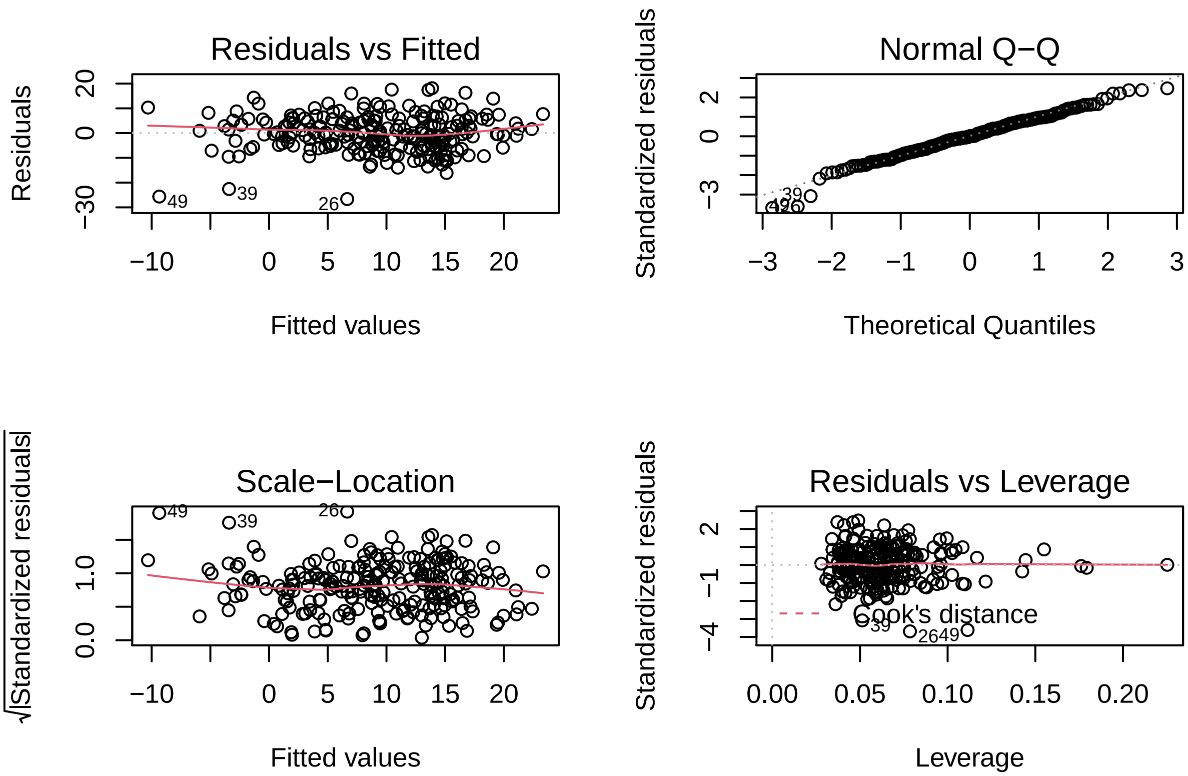


Figure C8. Results residuals from multivariate test model of the improvement in number and each graph is checking assumptions.

```
##                      GVIF Df GVIF^(1/(2*Df))
## site             1.658197  4        1.065257
## viewcat          1.642134  3        1.086179
## age              1.173242  1        1.083163
## encour           1.314660  1        1.146586
## prenumb:prelet   1.447865  1        1.203273
## prelet:diffbody  1.141987  1        1.068638
## prelet:diffclasf 5.478419  1        2.340602
## diffclasf:preform 5.431156  1        2.330484
```

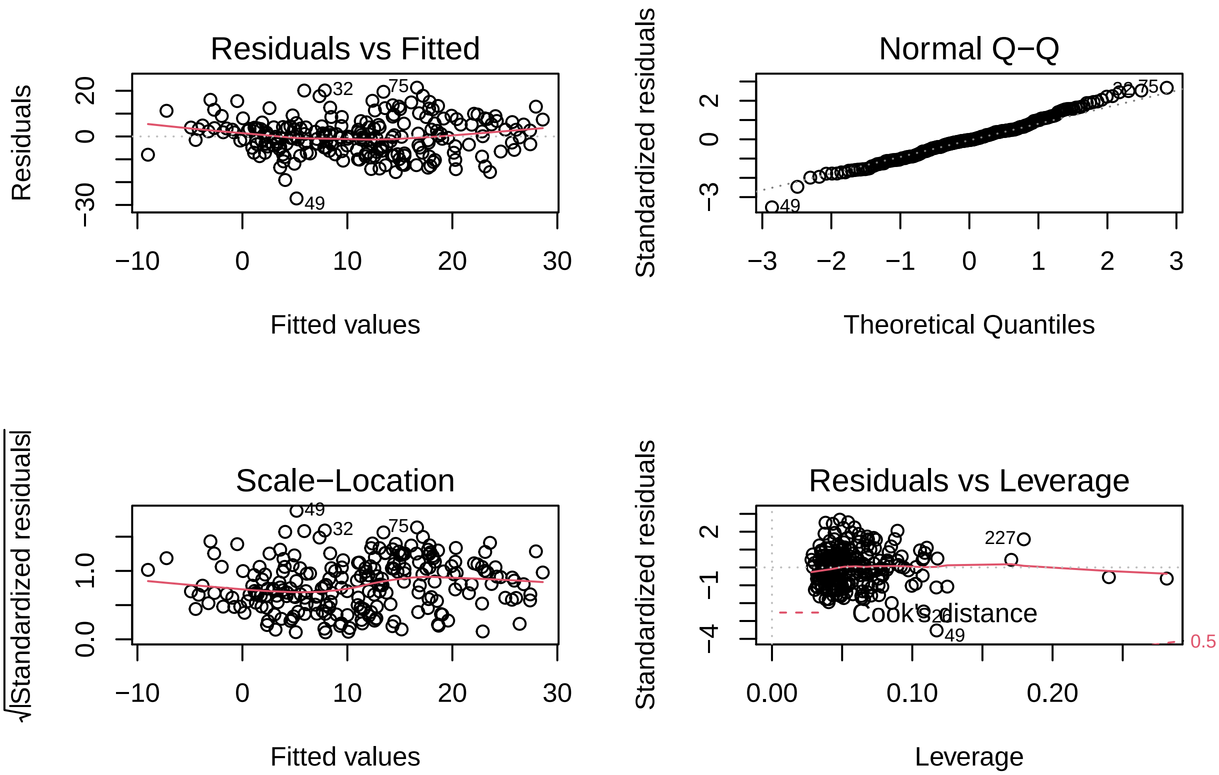


Figure C9. Results residuals from multivariate test model of the improvement in letter and each graph is checking assumptions.

```
##              GVIF Df GVIF^(1/(2*Df))
## viewcat 1.787581  3        1.101652
## age     1.189734  1        1.090749
## site    1.722225  4        1.070314
## preform 1.592999  1        1.262141
## diffrel 1.195634  1        1.093451
## encour  1.298949  1        1.139714
```

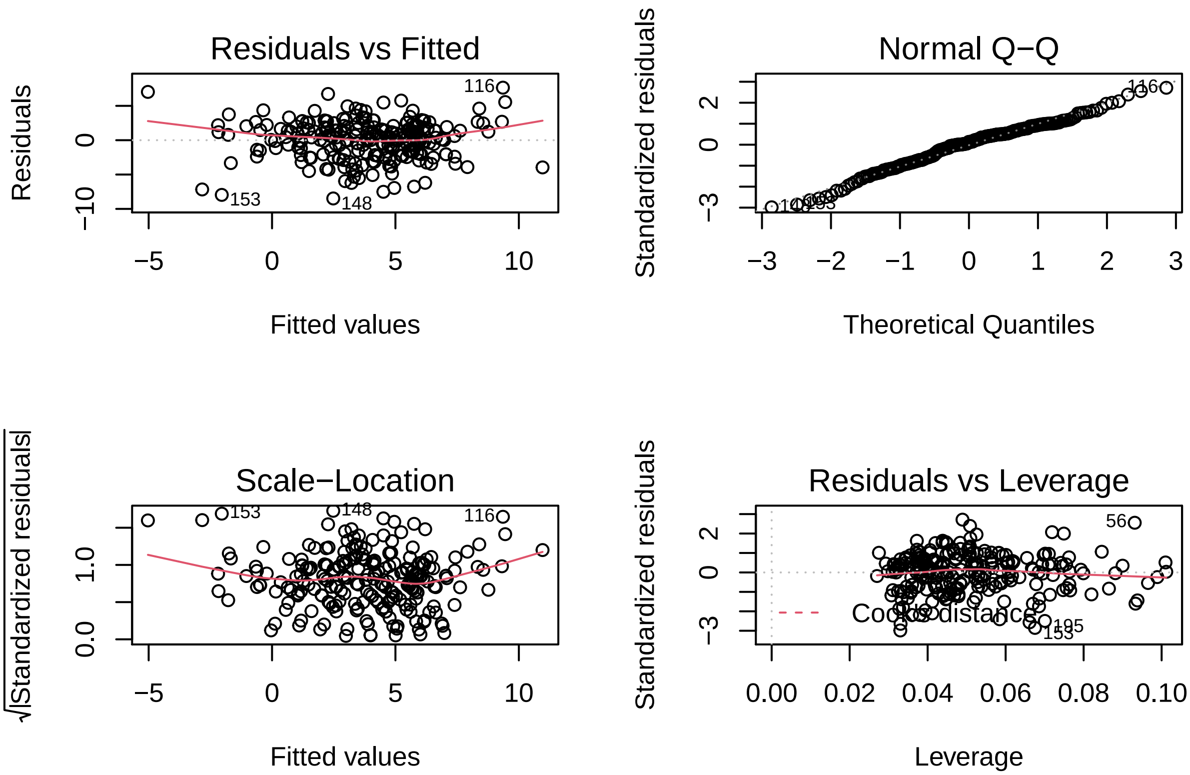


Figure C10. Results residuals from multivariate test model of the improvement in form and each graph is checking assumptions.

R Script

```
# clean up & set default chunk options
rm(list = ls())
knitr::opts_chunk$set(echo = FALSE)

# packages
library(tidyverse)
library(ggplot2)
library(mosaic)
library(table1)
library(readxl)
library(car)
library(stargazer)
library(tidyr)
library(MASS)
library(corrplot)
library(table1)
library(gtsummary)
library(interactions)
library(ggpubr)
# user-defined functions

# inputs
sesame = read_xlsx("sesame.xlsx")

#data structure
str(sesame)
#Missing Value
sum(is.na(sesame))
#Summary on Categorical values
sesame_cat <- sesame %>%
  select(site,sex,agecat,viewcat,setting,viewenc)
sesame_cat$agecat = sesame_cat$agecat+1 #1 below 51 2 above 51
tbl_summary(sesame_cat, by=site) %>%
  add_n()%>%
  add_stat_label()%>%
  modify_spanning_header(starts_with("stat_") ~ "**Raw Data Five Populations were of Interest(1,2,3,4,5
#Summary on All exam values
rawT<-table1( ~ prebody+postbody+prelet+postlet+preform+postform+prenumb+postnumb+prerelat+postrelat+pre

#Numerical Data Pre-processing
#omit the outlier than exceed the maximum score
sesame2 <- sesame
sesame2 = sesame2[sesame2$postbody <= 32,]
sesame2 = sesame2[sesame2$postlet <= 58,]
sesame2 = sesame2[sesame2$postrelat <= 17,]
#calculate the difference between pre-tests and post-tests
sesame2$diffbody = sesame2$postbody - sesame2$prebody
sesame2$difflet = sesame2$postlet - sesame2$prelet
sesame2$diffform = sesame2$postform - sesame2$preform
sesame2$diffnumb = sesame2$postnumb - sesame2$prenumb
sesame2$diffrel = sesame2$postrelat - sesame2$prerelat
sesame2$diffclasf = sesame2$postclasf - sesame2$preclasf
```

```
#Categorical Data Pre-processing
#change setting to home(0) and school(1); male(0), female(1); encourage(0), not (1)
sesame2$setting[sesame2$setting == 1] = 0 #home
sesame2$setting[sesame2$setting == 2] = 1 #school
sesame2$sex[sesame2$sex == 1] = 0 #male
sesame2$sex[sesame2$sex == 2] = 1 #female
sesame2$viewenc[sesame2$viewenc == 1] = 0 #encourage
sesame2$viewenc[sesame2$viewenc == 2] = 1 #not encourage

#Factor all categorical variables
sesame2$site <- as.factor(sesame2$site)
sesame2$sex <- as.factor(sesame2$sex)
sesame2$agecat <- as.factor(sesame2$agecat)
sesame2$viewcat <- as.factor(sesame2$viewcat)
sesame2$setting <- as.factor(sesame2$setting)
sesame2$viewenc <- as.factor(sesame2$viewenc)
#plot explanatory variable distribution
s<-ggplot(sesame2, aes(x = site)) + geom_bar()
se<-ggplot(sesame2, aes(x = sex)) + geom_bar()
a<-ggplot(sesame2, aes(x = age)) + geom_bar()
v<-ggplot(sesame2, aes(x = viewcat)) + geom_bar()
set<-ggplot(sesame2, aes(x = setting)) + geom_bar()
ve<-ggplot(sesame2, aes(x = viewenc)) + geom_bar()
ggarrange(s,se,a,v,set,ve)
sesame_cat2 <- sesame2 %>%
  select(site,sex,agecat,viewcat,setting,viewenc)
tbl_summary(sesame_cat2, by=site) %>%
  add_n()%>%
  add_stat_label()%>%
  modify_spanning_header(starts_with("stat_") ~ "**Five Populations were of Interest(1,2,3,4,5)**")
#summary table on test difference
table1::label(sesame2$diffbody) <- "body parts"
table1::label(sesame2$difflet) <- "letters"
table1::label(sesame2$diffform) <- "forms"
table1::label(sesame2$diffnumb) <- "numbers"
table1::label(sesame2$diffrel) <- "relations"
table1::label(sesame2$diffclasf) <- "classification"
afterT<-table1(~diffbody + difflet + diffform + diffnumb + diffrel + diffclasf | site, data = sesame2, 
# improvement v.s viewcat
par(mfrow=c(1,3))
# letter v.s viewcat
boxplot(sesame2$difflet~sesame2$viewcat, main= "Boxplot improvement on letter", xlab = "Viewing Frequenc
# number v.s viewcat
boxplot(sesame2$diffnumb ~ sesame2$viewcat, main= "Boxplot improvement on number", xlab = "Viewing Frequ
# form v.s viewcat
boxplot(sesame2$diffform ~ sesame2$viewcat, main= "Boxplot improvement on form", xlab = "Viewing Freque
# improvement vs.age
par(mfrow=c(2,2))
# letter v.s age
boxplot(sesame2$difflet ~ sesame2$age, main= "Boxplot improvement on letter", xlab = "Age", ylab = "dif
# form v.s age
boxplot(sesame2$diffform ~ sesame2$age, main= "Boxplot improvement on form", xlab = "Age", ylab = "diff
# number v.s age
```

```
boxplot(sesame2$diffnumb ~ sesame2$age, main= "Boxplot improvement on number", xlab = "Age", ylab = "di
# improvement v.s site
par(mfrow=c(1,3))
# letter v.s site
boxplot(sesame2$difflet ~ sesame2$site, main= "Boxplot improvement on letter", xlab = "site", ylab = "di
# form v.s site
boxplot(sesame2$diffform ~ sesame2$site, main= "Boxplot improvement on form", xlab = "site", ylab = "di
# number v.s site
boxplot(sesame2$diffnumb ~ sesame2$site, main= "Boxplot improvement on number", xlab = "site", ylab = "
# improvement v.s peabody
par(mfrow=c(2,2))
# letter v.s peabody
boxplot(sesame2$difflet ~ sesame2$peabody, main= "Boxplot improvement on letter", xlab = "peabody", yla
# form v.s peabody
boxplot(sesame2$diffform ~ sesame2$peabody, main= "Boxplot improvement on form", xlab = "peabody", ylab
# number v.s peabody
boxplot(sesame2$diffnumb ~ sesame2$peabody, main= "Boxplot improvement on number", xlab = "peabody", yl
# Rest not important in my view (encourage, setting, sex)
# improvement v.s encourage
par(mfrow=c(3,3))
# letter v.s encour
boxplot(sesame2$difflet ~ sesame2$encour, main= "Boxplot improvement on letter", xlab = "encour", ylab
# form v.s encour
boxplot(sesame2$diffform ~ sesame2$encour, main= "Boxplot improvement on form", xlab = "encour", ylab =
# number v.s encour
boxplot(sesame2$diffnumb ~ sesame2$encour, main= "Boxplot improvement on number", xlab = "encour", ylab
# v.s setting
# letter v.s setting
boxplot(sesame2$difflet ~ sesame2$setting, main= "Boxplot improvement on letter", xlab = "setting", yla
# form v.s setting
boxplot(sesame2$diffform ~ sesame2$setting, main= "Boxplot improvement on form", xlab = "setting", ylab
# number v.s setting
boxplot(sesame2$diffnumb ~ sesame2$setting, main= "Boxplot improvement on number", xlab = "setting", yl
# v.s sex
# letter v.s sex
boxplot(sesame2$difflet ~ sesame2$sex, main= "Boxplot improvement on letter", xlab = "sex", ylab = "dif
# form v.s sex
boxplot(sesame2$diffform ~ sesame2$sex, main= "Boxplot improvement on form", xlab = "sex", ylab = "diff
# number v.s sex
boxplot(sesame2$diffnumb ~ sesame2$sex, main= "Boxplot improvement on number", xlab = "sex", ylab = "di
sesame3 <- sesame2
sesame3 <- sapply(sesame3, as.numeric)
sesame.cor <- cor(sesame3)
corrplot(sesame.cor, order = "hclust", addrect = 2, tl.col = "black", title = "Correlations Plot betwee
# Number Test Modeling
mod_num <- lm(diffnumb ~., data = sesame2)
summary(mod_num)
mod1_num <- lm(diffnumb ~ (factor(site) + factor(sex) + age + factor(viewcat) + factor(setting) + facto
summary(mod1_num)
step.model_num <- stepAIC(mod1_num, direction = "both",
                      trace = FALSE)
summary(step.model_num)
mod_num1 <- lm(diffnumb ~ site + sex + viewcat + prenumb + preform + diffbody + diffclasf + diffrel + s
```

```
summary(mod_num1)

mod_num2 <- lm(diffnumb ~ prenumb + diffclasf + diffbody + diffrel + diffbody:age + site:age + site:prel
summary(mod_num2) #site,encour,age,viewcat,
# Letter Test Modeling
mod_let <- lm(difflet ~ (factor(site) + age + factor(sex) + factor(viewcat) + factor(setting) + factor(
summary(mod_let)
step.model_let <- stepAIC(mod_let, direction = "both",
                      trace = FALSE)
summary(step.model_let)

mod_let1 <- lm(difflet ~ site + age + viewcat + prenumb + preform + diffbody + diffclasf + diffrel + pre
summary(mod_let1)

mod_let2 <- lm(difflet ~ site + viewcat + age + prenumb + encour + prenumb:prelet + diffbody:prelet + di
summary(mod_let2)
# Form Test Modeling
mod_form <- lm(diffform ~ (factor(site) + age + factor(sex) + factor(viewcat) + factor(setting) + factor
summary(mod_form)

step.model_form <- stepAIC(mod_form, direction = "both",
                      trace = FALSE)
summary(step.model_form)

mod_form1 <- lm(diffform ~ site + sex + viewcat + peabody + preform + diffbody + diffrel + age:viewcat +
summary(mod_form1)

mod_form2 <- lm(diffform ~ site + sex + age + encour + viewcat + preform + diffrel + diffrel:prelet, dat
summary(mod_form2)
vif(num_fit)
par(mfrow = c(2,2))
plot(num_fit)
vif(let_fit)
par(mfrow = c(2,2))
plot(let_fit)
vif(form_fit)
par(mfrow = c(2,2))
plot(form_fit)
```