How school type, language, special education services, family income, and parental education level influence OSSLT first attempt results in Ontario schools

STA302 Final Project Part 1

Xuanle Zhou Luhan Wang Junyi Hou

April 7, 2025

1 Introduction

The Ontario Secondary School English Literacy Test (OSSLT) is mandatory for high school graduation in Ontario, therefore English language learning is a significant focus for both parents and students. This paper aims to investigate how school type, language, special education services, family income, and parental education level influence OSSLT first attempt results in Ontario schools. Zhang et al. (2020) found that family income and parental education level significantly contribute to a student's academic success. Their study was conducted in China, and revealed that higher family income and more advanced parental education are correlated with better student performance. This supports and shapes our hypothesis that students with higher family income and parental education level will perform better on the OSSLT. Bernhofer and Tonin (2022) showed that students perform better when taught in their first language, which questions if English language school students will perform better on the OSSLT compared to those in non-English language schools. Lastly, Aseery (2024) explored how technology and multimedia elements in religious education classes could enhance English language learning. Aseery's findings suggest that multimedia tools in religious education classes improve student engagement and motivation, which enhances learning outcomes. We would expect schools supplying these technologies in 2025. Therefore, we hypothesize that religious schools will have higher OSSLT pass rates.

While Zhang et al. (2020) concluded that higher income and parental education level lead to higher achievement, there are exceptions, as many successful individuals come from lower-income backgrounds. We also expect that students receiving special education services may

Table 1: OSSLT First Attempt Pass Rate Descriptive Statistics

	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
OSSLT_First_Attempt_	737	82.45	11.93	85	83.92	10.38	0	100	100	-1.93	6.93	0.44

perform worse on the OSSLT due to specific learning disabilities, despite receiving accommodations. This research question fits well with the concept of multiple linear regression, which examines how multiple predictor variables collaboratively influence a response variable. Therefore, we have selected multiple linear regression as our analysis method. Since the main goal is to observe patterns between variables, this model will focus on interpretability.

This research will benefit those seeking an accurate analysis of the factors that influence English learning outcomes, particularly in the context of the OSSLT. The response variable, OSSLT results, serves as an effective measure of students' English proficiency, as it is both a pass/fail test and provides continuous data.

2 Data description

The dataset, available on the Ontario Data Catalogue (Ontario 2024b), provides insights into schools in Ontario, supporting policy-making, and educational research. This study repurposes it to investigate and predict the OSSLT first-attempt pass rate. Data were collected from schools, school boards, EQAO, and Statistics Canada through online forms, surveys, phone interviews, and in-person visits, then compiled by Ontario Data Catalogue (Ontario 2024a).

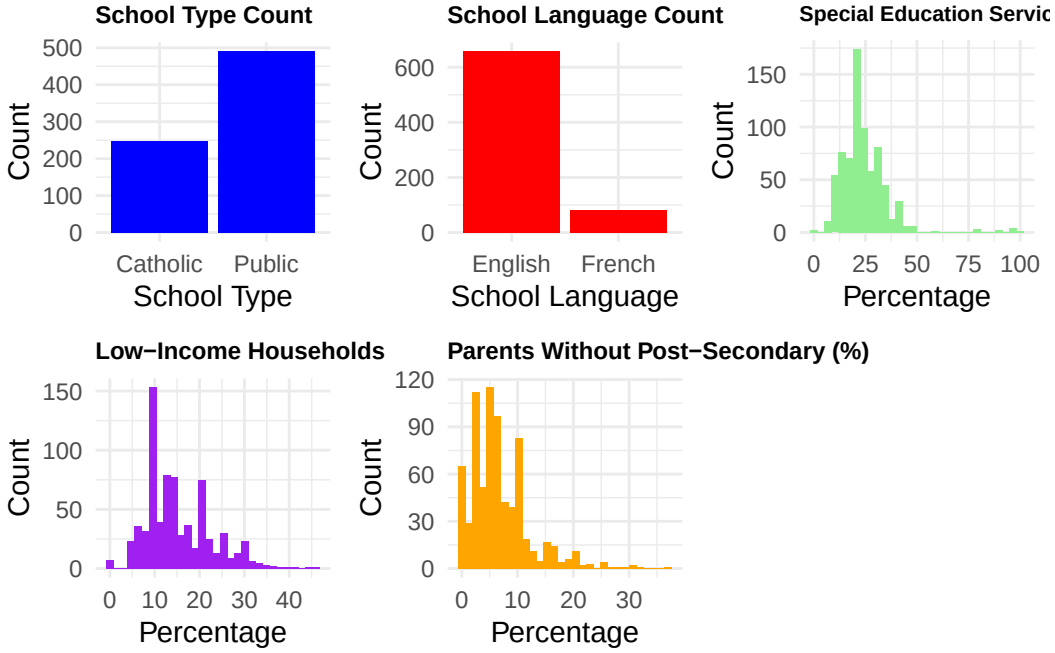
The `OSSLT_First_Attempt_PassRate`, the response variable, measures the percentage of students passing Ontario Secondary School Literacy Test on their first attempt, ranging from 0 to 100. The mean of 82.45 and median of 85 indicate high pass rates. The dataset originally had 4,926 observations, reduced to 737 after cleaning, ensuring statistical reliability. Despite being bounded, the pass rate is continuous, suitable for linear regression.

```
                              n  mean    sd median trimmed   mad min max range
OSSLT_First_Attempt_PassRate 737 82.45 11.93     85   83.92 10.38   0 100   100
                             skew kurtosis   se
OSSLT_First_Attempt_PassRate -1.93     6.93 0.44
```

School Type is categorical, with two types: Catholic and Public. Most schools are public. Cheema (2024) noted, private schools generally outperform public schools in literacy. We expect Catholic schools to have higher OSSLT pass rates due to structured curriculum and discipline.

School Language is binary, English or French. Most schools operate in English, which is expected to correlate with higher OSSLT pass rates.

Table 2: Histograms for Selected Predictors



Students receiving special education services often exhibit lower literacy achievement and slower progress, as noted by Vaughn and Wanzek (2014). Our model aims to capture this pattern. The mean of this predictor variable is 24.07%, with a median of 22%, includes outliers where 100% of students receive special education services.

The percentage of school-aged children in low-income households has a mean of 15.27% and skewness of 0.88, indicating some schools have significantly higher concentrations. As Nadeem, Akhtar, and Ahmad (2021) found, lower-income students often have lower literacy skills, which we expect to correlate with lower OSSLT pass rates.

The percentage of students whose parents lack post-secondary credentials averages 6.76%, with skewness of 1.56 and kurtosis of 3.73, suggesting a slight right skew. As Davis-Kean, Tighe, and Waters (2021) states, parental education influences children's academic success, making this a relevant predictor.

3 Preliminary results

```
Loading required package: carData
```

```
Attaching package: 'car'

The following object is masked from 'package:psych':

    logit

The following object is masked from 'package:dplyr':

    recode
```

	Coefficient	Standard_Error	t_Statistic	p_Value
(Intercept)	105.1756826	1.00518235	104.633435	0.000000e+00
School_TypePublic	-1.0689414	0.65175574	-1.640095	1.014156e-01
School_LanguageFrench	5.4585354	0.98401923	5.547184	4.059800e-08
Special_Ed_Pct	-0.6210042	0.02613604	-23.760460	6.428134e-93
Low_Income_Pct	-0.2957966	0.04835594	-6.117068	1.552365e-09
No_Parent_Degree_Pct	-0.4645510	0.06539550	-7.103716	2.886850e-12

3.1 Residual Analysis

3.1.1 Linear Models Assumptions:

1. **Linearity**

$$E(Y_i|X = \mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

2. **Constant Error Variance (Homoscedasticity)**

$$Var(Y_i|X = \mathbf{x}_i) = \sigma^2$$

3. **Uncorrelated and Normal Errors**

$$Cov(e_i, e_j) = 0 \text{ for } i \neq j \text{ and } e_i \sim N(0, \sigma^2)$$

3.1.2 Assumption Check

```
function ()
{
    for (fun in getHook("before.plot.new")) {
        if (is.character(fun))
            fun <- get(fun)
        try(fun())
    }
    .External2(C_plot_new)
    grDevices:::recordPalette()
    for (fun in getHook("plot.new")) {
        if (is.character(fun))
            fun <- get(fun)
        try(fun())
    }
    invisible()
}
<bytecode: 0x1300f8f30>
<environment: namespace:graphics>
```

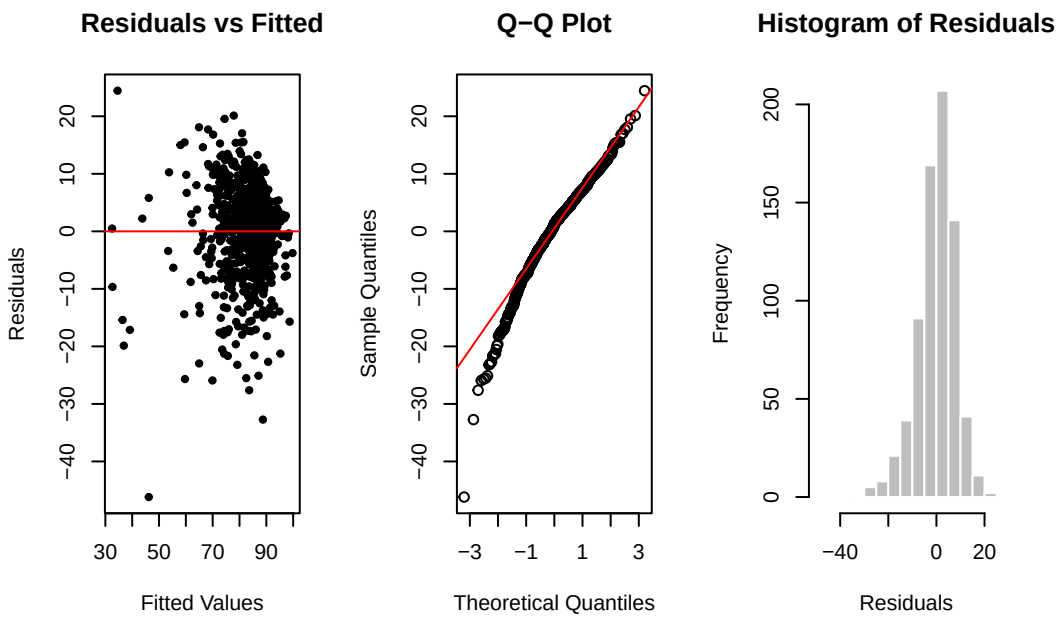


Figure 1: Residual Plots

1. **Linearity & Homoscedasticity:** The residuals vs. fitted plot shows no clear pattern, suggesting linearity. Slight heteroscedasticity is observed.
2. **Normality:** The Q-Q plot and the histogram of residuals suggest residuals are approximately normal, though slight deviations exist at the left tail.
3. **Independence:** No evident pattern in the residual plot suggests residuals are independent.

3.2 Model Interpretation & Discussion

3.2.1 Key Findings and interpretation

- The **intercept (105.18)** represents the estimated pass rate for a **Catholic, English-language school with 0% special education, 0% low-income students, and 0% students whose parents have no degree**. This provides a reference point for understanding the model's predictions.
- **School Language (French vs. English)** and the three numeric variables (**Special_Ed_Pct, Low_Income_Pct, No_Parent_Degree_Pct**) are strongly associated with the **OSSLT pass rate**.
- **School Type (Public vs. Catholic)** does not show a statistically significant difference in pass rate in this model.
- Higher proportions of **special education students, low-income students, and students whose parents have no degree** are each associated with a **lower pass rate**.
- Conversely, being a **French-language school** is associated with a **higher pass rate** relative to the English.
- The model explains **54% of the variation in pass rates**, which is reasonable for educational data, suggesting these variables collectively have a substantial but not complete ability to predict pass rates.

3.2.2 Comparison to Literature

Our findings align with prior research while offering insights specific to Ontario:

- **Family Income & Parental Education:** Consistent with Zhang et al. (2020), our results confirm that higher family income and parental education correlate with better OSSLT pass rates.
- **School Language:** Contrary to Bernhofer and Tonin (2022), our study shows French-language schools had higher OSSLT pass rates than English-language schools, indicating other factors like curriculum or funding may play a role. Further investigation is needed.

Table 3: Descriptive Statistics for Selected Predictors

	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
School_Type*	737	1.66	0.47	2	1.71	0.00	1	2	1	-0.70	-1.52	0.02
School_Language*	737	1.11	0.31	1	1.01	0.00	1	2	1	2.51	4.31	0.01
Special_Ed_Pct	737	24.07	11.53	22	22.91	8.90	0	100	100	2.72	13.46	0.42
Low_Income_Pct	737	15.27	7.30	13	14.58	5.93	0	46	46	0.88	0.66	0.27
No_Parent_Degree_Pct	737	6.76	5.42	5	6.06	4.45	0	37	37	1.56	3.73	0.20

- **Special Education:** Higher proportions of special education students negatively impact OSSLT success, aligning with expectations.
- **School Type:** No significant difference was found between public and Catholic schools, despite Aseery (2024) suggesting that religious schools may benefit from enhanced multimedia learning tools.

```
                      n  mean    sd median trimmed  mad min max range  skew
School_Type*        737  1.66  0.47      2    1.71 0.00   1   2     1 -0.70
School_Language*    737  1.11  0.31      1    1.01 0.00   1   2     1  2.51
Special_Ed_Pct      737 24.07 11.53     22   22.91 8.90   0 100   100  2.72
Low_Income_Pct      737 15.27  7.30     13   14.58 5.93   0  46    46  0.88
No_Parent_Degree_Pct 737  6.76  5.42      5    6.06 4.45   0  37    37  1.56
                     kurtosis   se
School_Type*            -1.52 0.02
School_Language*         4.31 0.01
Special_Ed_Pct          13.46 0.42
Low_Income_Pct           0.66 0.27
No_Parent_Degree_Pct     3.73 0.20
```

3.2.3 Obtain Box Cox Suggestion

```
Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select

[1] 2
```

3.2.4 Response Variable Transformation Preview with QQ

- We got 2 as Box Cox lambda, Therefore we Try Y^2 Transformation, we also tried many other transformations such as 1/Y, and decided to show Log and Square Root and they are the most improved Models.

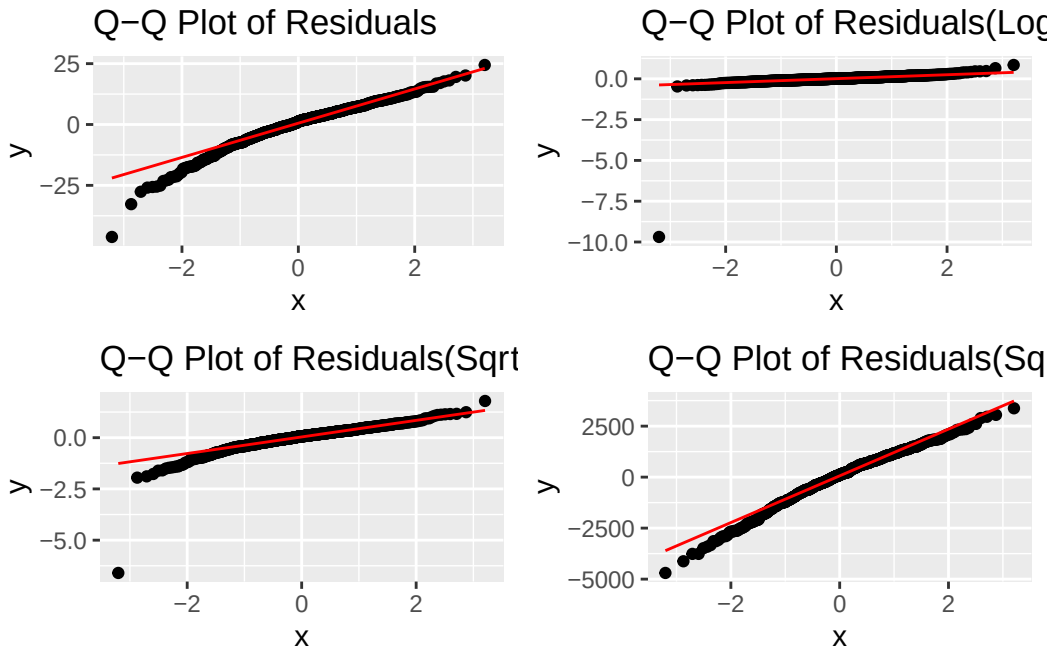


Figure 2: Residual Plots

3.2.5 Response Variable Transformation Preview with Residuals

- Square root looks the best for residual plot

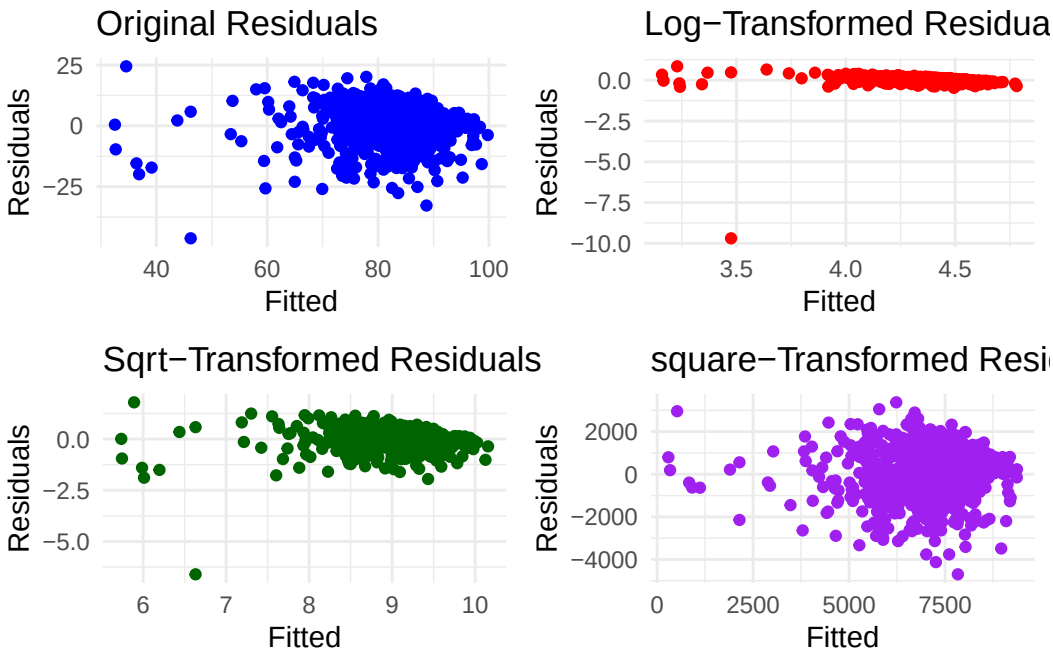


Figure 3: Residual Plots

3.2.6 Response Variable Transformation Decision

- We eliminate Square since the model performs worse in every aspect. We need to choose one between Log and Square Root Transformation that both improve from the original model. We ended up with Square root since R squared for square root is significantly higher to Log. The difference in AIC, BIC and AICc is not significant enough to replace R^2.

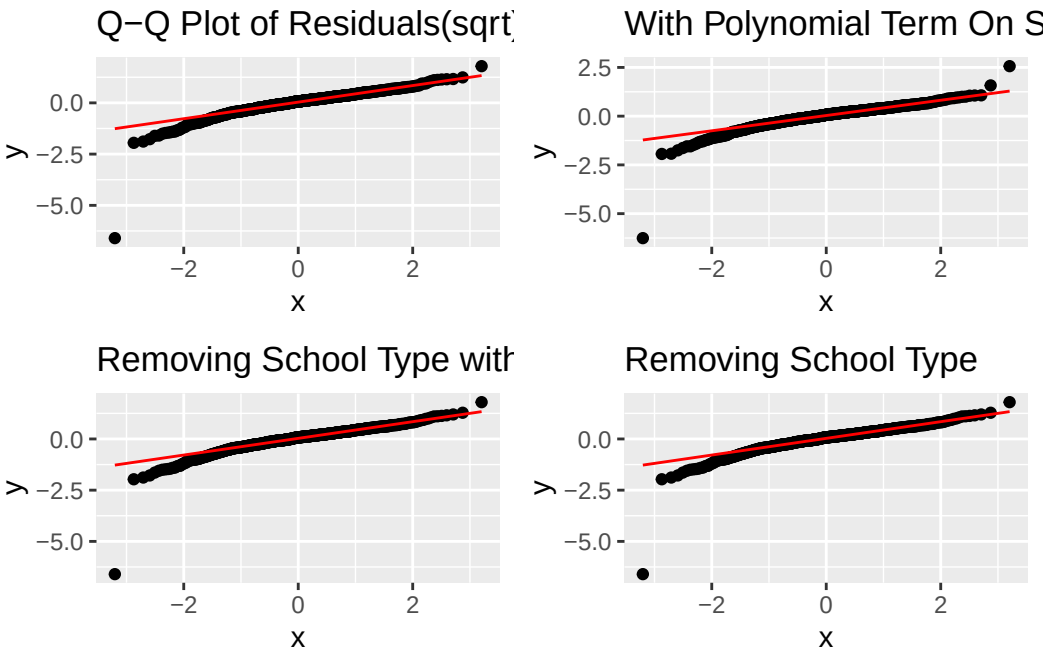
	Model	R_Squared	AIC	BIC	AICc
1	Original	0.5415454	5179.0167	5211.2348	5179.1703
2	Log-Transformed	0.1998555	687.2028	719.4209	687.3564
3	Sqrt-Transformed	0.5265497	1168.6607	1200.8788	1168.8143
4	Square	0.5056889	12565.1800	12597.3981	12565.3337

3.2.7 Y transformed model summary, VIF and Confidence Interval

- Confidence Interval contains 0 and p value >0.05 for School_Type, Suggest dropping this predictor.

```
Call:
lm(formula = OSSLT_Sqrt ~ School_Type + School_Language + Special_Ed_Pct +
    Low_Income_Pct + No_Parent_Degree_Pct, data = df_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6003 -0.2370  0.0663  0.3083  1.7902

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           10.511986   0.066168 158.869  < 2e-16 ***
School_TypePublic     -0.052970   0.042903  -1.235    0.217
School_LanguageFrench  0.298159   0.064775   4.603 4.91e-06 ***
Special_Ed_Pct        -0.041371   0.001720 -24.047  < 2e-16 ***
Low_Income_Pct        -0.019850   0.003183  -6.236 7.59e-10 ***
No_Parent_Degree_Pct  -0.024074   0.004305  -5.592 3.16e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5318 on 731 degrees of freedom
Multiple R-squared:  0.5298,    Adjusted R-squared:  0.5265
F-statistic: 164.7 on 5 and 731 DF,  p-value: < 2.2e-16


         School_Type      School_Language       Special_Ed_Pct
            1.068840             1.058065             1.023733
      Low_Income_Pct No_Parent_Degree_Pct
            1.406403             1.416469


                           2.5 %      97.5 %
(Intercept)           10.38208456 10.64188781
School_TypePublic     -0.13719718  0.03125808
School_LanguageFrench  0.17099240  0.42532576
Special_Ed_Pct        -0.04474875 -0.03799353
Low_Income_Pct        -0.02609894 -0.01360068
No_Parent_Degree_Pct  -0.03252536 -0.01562299
```

3.2.8 X Transformation QQ Preview



3.2.9 X Transformation Residual Plots

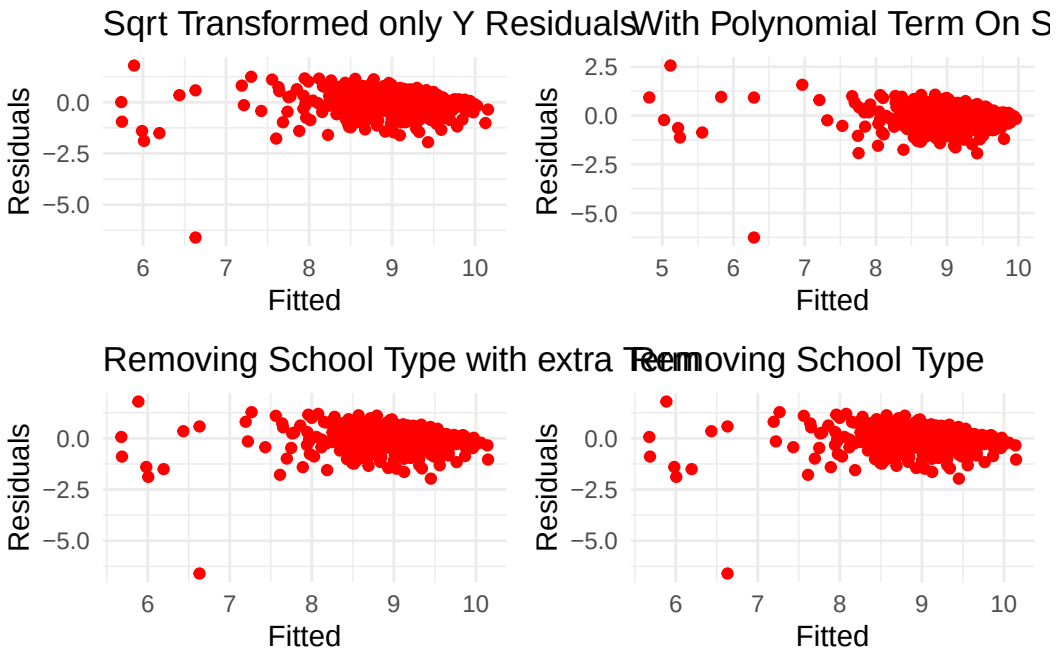


Figure 4: Residual Plots

3.2.10 Predictor Transformation Decision

- We decide to choose Model 3: X Transform with one less Predictor, since the R^2, AIC, BIC, AICc improved from the original sqrt-Transformed. Model 2 is slightly better but it contains an extra predictor.

```
                                Model R_Squared       AIC      BIC     AICc
1                   Sqrt-Transformed 0.5265497 -924.8547 1200.879 1168.814
2                        X Transform 0.5458389 -954.5191 1175.817 1139.194
3 X Transform with one less Predictor 0.5262106 -925.3194 1195.811 1168.311
4                  One Less Predictor 0.5262106 -925.3194 1195.811 1168.311
```

3.2.11 Anova Table for X Transformation

- Finally, we can conclude the final model with 5 Predictors: School Language, Special Education Percentage, Special Education Percentage Squared, Lower Income background and Parents with no Education. Because RSS improved.

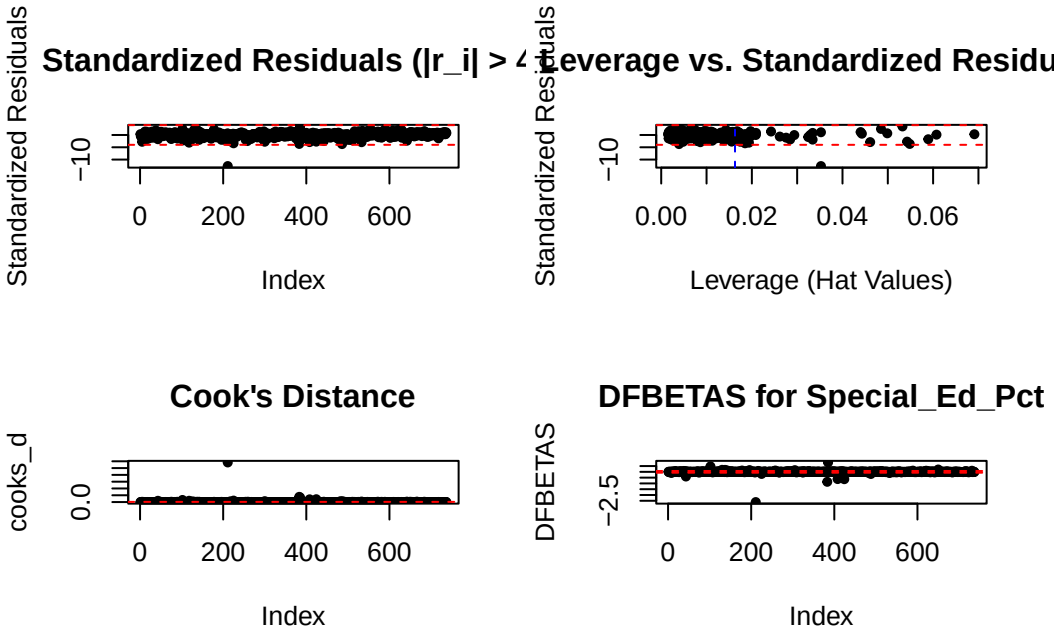
```
Analysis of Variance Table

Model 1: OSSLT_Sqrt ~ School_Type + School_Language + Special_Ed_Pct +
    Sp_Ed_Transform + Low_Income_Transform + No_Edu_Transform
Model 2: OSSLT_Sqrt ~ School_Type + School_Language + Special_Ed_Pct +
    Low_Income_Pct + No_Parent_Degree_Pct
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    730 198.04
2    731 206.73 -1   -8.6939 32.047 2.163e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Analysis of Variance Table

Model 1: OSSLT_Sqrt ~ School_Language + Special_Ed_Pct + Low_Income_Transform +
    No_Edu_Transform
Model 2: OSSLT_Sqrt ~ School_Type + School_Language + Special_Ed_Pct +
    Low_Income_Pct + No_Parent_Degree_Pct
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    732 207.16
2    731 206.73  1   0.43109 1.5243 0.2174


Analysis of Variance Table

Model 1: OSSLT_Sqrt ~ School_Language + Special_Ed_Pct + Low_Income_Transform +
    No_Edu_Transform
Model 2: OSSLT_Sqrt ~ School_Type + School_Language + Special_Ed_Pct +
    Low_Income_Pct + No_Parent_Degree_Pct
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    732 207.16
2    731 206.73  1   0.43109 1.5243 0.2174
```

3.2.12 Outlier Detection and Removal

- 211,385,102,118,225,484,533. These columns appear under several different tests, so we tried to fit the model after removing these.

```
=== Outliers ===


Standardized residuals (|r_i| > 4): 211
```

```
Studentized residuals (|rstudent| > 3): 118 211 225 383 385 486

=== Influence Metrics ===

High Leverage (Hat > 0.016 ): 43 45 47 58 59 61 63 71 73 74 86 101 102 106 118 122 125 137 2

Influential (Cook's D > 0.005 ): 5 43 47 75 90 102 104 106 118 125 130 137 142 178 196 211 2

High DFFITS (> 0.165 ): 5 43 47 75 90 102 104 106 118 125 130 137 142 178 196 211 217 225 254

High DFBETAS (> 0.074 ): 5 8 18 38 43 47 50 51 57 58 62 67 71 72 75 77 81 82 84 87 89 90 94
```



- Cleaned Model Summary(Without outliers) and Confidence Interval

```
Call:
lm(formula = OSSLT_Sqrt ~ School_Language + Special_Ed_Pct +
    Low_Income_Transform + No_Edu_Transform, data = df_clean[-c(211,
    385, 102, 118, 225, 484, 533), ])

Residuals:
```

```
     Min       1Q   Median       3Q      Max
-2.11062 -0.23415  0.07156  0.29064  1.10847

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           10.391501   0.054410 190.986  < 2e-16 ***
School_LanguageFrench  0.312840   0.054370   5.754 1.29e-08 ***
Special_Ed_Pct        -0.039004   0.001542 -25.291  < 2e-16 ***
Low_Income_Transform  -0.015892   0.002729  -5.824 8.65e-09 ***
No_Edu_Transform      -0.026883   0.003656  -7.353 5.26e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4504 on 725 degrees of freedom
Multiple R-squared:  0.5607,    Adjusted R-squared:  0.5583
F-statistic: 231.3 on 4 and 725 DF,  p-value: < 2.2e-16


                            2.5 %      97.5 %
(Intercept)           10.28468151 10.49832065
School_LanguageFrench  0.20609878  0.41958045
Special_Ed_Pct        -0.04203181 -0.03597634
Low_Income_Transform  -0.02124993 -0.01053486
No_Edu_Transform      -0.03406083 -0.01970472
```

3.2.13 Assess Model Performance after removing outlier

- Hooray, AIC, BIC, AICc, R^2 all improved significant enough to conclude the removing outlier process.

```
             Model R_Squared        AIC       BIC      AICc
1 Sqrt-Transformed 0.5262106  -925.3194 1195.8115 1168.3110
2 Outliers Removed 0.5582506 -1159.3910  941.8175  914.3755
```

3.2.14 QQ and Residual comparsion

- Both looks Better

Figure 5: Residual Plots

4 Final model inference and results

Table 4: Regression Coefficients with 95% Confidence Intervals

Predictor	Estimate	Std. Error	t value	p-value	Lower 95% CI	Upper 95% CI
(Intercept)	10.3915	0.0544	190.99	< 2e-16	10.2847	10.4983
School Language - French	0.3128	0.0544	5.75	1.29e-08	0.2061	0.4196
% in Special Education	-0.0390	0.0015	-25.29	< 2e-16	-0.0420	-0.0360
% in Low Income	-0.0159	0.0027	-5.82	8.65e-09	-0.0212	-0.0105
% in No Educated Parent	-0.0269	0.0037	-7.35	5.26e-13	-0.0341	-0.0197

4.1 Model Interpretation

The regression results presented in Table 4 provide meaningful insight into how school and family level factors influence OSSLT first attempt success rates in Ontario. Notably, the language of instruction emerges as a significant predictor: schools offering instruction in French are associated with higher OSSLT performance, holding all other variables in the model constant. Specifically, French language schools are predicted to have a 0.31 unit increase in the square root of the OSSLT first attempt pass rate compared to English language schools. The

95% confidence interval for this estimate ranges from 0.2061 to 0.4196, indicating a consistently positive effect that is statistically significant.

In contrast, three key indicators of socioeconomic and educational disadvantage are all significantly associated with lower OSSLT performance: the percentage of students receiving special education services, the percentage of school aged children living in low income households, and the percentage of students whose parents have no degree, diploma, or certificate. For each one percentage point increase in students receiving special education, there is an estimated 0.039 unit decrease in the square root of the OSSLT pass rate, with a 95% confidence interval ranging from -0.0420 to -0.0360. Likewise, each additional percentage point of students from low income households is associated with a 0.0159 unit decrease, with the confidence interval ranging from -0.0212 to -0.0105. Finally, each percentage point increase in students whose parents lack post secondary education corresponds to a 0.0269 unit decrease in the transformed OSSLT outcome, with a confidence interval between -0.0341 and -0.0197. The narrow confidence intervals for all predictors suggest that the estimated effects are both precise and robust.

4.2 Comparing with Literature

The findings from our final regression model align closely with much of the existing literature on factors influencing student literacy outcomes. Consistent with Zhang et al. (2020), we found that both family income and parental education level are significant predictors of OSSLT success: schools with higher percentages of students from low-income households and students whose parents lack post-secondary education showed notably lower OSSLT pass rates. This supports the broader claim that socioeconomic status plays a critical role in shaping educational achievement. Similarly, our results reinforce the observations of Vaughn and Wanzek (2014), as schools with higher proportions of students receiving special education services were significantly associated with lower OSSLT outcomes, likely due to the academic challenges these students face, even with accommodations.

However, our results diverge from the expectation presented by Bernhofer and Tonin (2022), who suggest students perform better when taught in their first language. In our analysis, schools offering instruction in French had significantly higher OSSLT performance, even though the OSSLT is administered in English. This suggests that French language instruction may be associated with school environments or educational practices that contribute positively to student literacy, despite the language difference. It may also reflect broader institutional or cultural differences between French and English schools in Ontario that warrant further exploration, such as school funding models, curriculum focus, or community engagement.

Our hypothesis regarding school type was not supported in the final model. School type, which identifies whether a school is Catholic or Public, was excluded due to a lack of statistical significance as discussed above. This outcome contrasts with the findings of Cheema (2024), who reported that private schools tend to outperform public schools in literacy achievement. While

we expected Catholic schools to demonstrate higher OSSLT performance due to structured curricula or the potential influence of religious education resources, our model did not find a meaningful difference once other variables were accounted for. This suggests that variation in OSSLT performance across schools is more strongly explained by socioeconomic and instructional factors than by school type alone.

Table 5: Model Fit Statistics

Metric	Value
R-squared	0.5607
Adjusted R-squared	0.5583
AIC	914.2593
BIC	941.8175
AICc	914.3755
Residual Std. Error	0.4504

4.3 Model Performance Assessment

The performance of the final multiple linear regression model, as shown in Table 5, can be evaluated using several statistical metrics that assess both goodness of fit and model parsimony. One of the most interpretable metrics is the R-squared value, which in this model is 0.5607. This indicates that approximately 56.1% of the variation in the square root of the OSSLT first-attempt pass rate is explained by the predictors included in the model. In the context of educational research, where student performance can be influenced by many unmeasured social, psychological, and institutional factors, an R-squared value above 0.5 is considered relatively strong. It suggests that the model captures a substantial portion of the meaningful variance across schools in Ontario.

The Adjusted R-squared value, which accounts for the number of predictors in the model, is 0.5583. While slightly lower than the unadjusted R-squared, this is expected and confirms that the included predictors contribute meaningfully to explaining the outcome without overfitting the data. The minimal difference between the two values suggests that the model achieves a good balance between explanatory power and complexity. This strengthens confidence that the model's performance is not artificially inflated by the number of predictors used.

Beyond explanatory power, model selection criteria such as the Akaike Information Criterion (AIC), the corrected AIC (AICc), and the Bayesian Information Criterion (BIC) provide important insights into model efficiency and generalizability. Both AIC and AICc are measures of model fit that penalize complexity, with lower values indicating better-fitting models. In our model, the AIC is 914.25 and the AICc is 914.36.

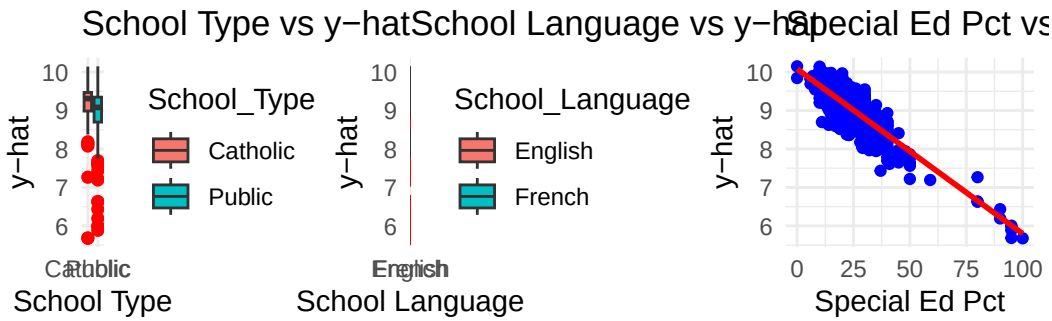
The distinction between AIC and AICc lies in their intended use: AICc is a bias-corrected version of AIC that is particularly useful when the sample size is small or when the number of

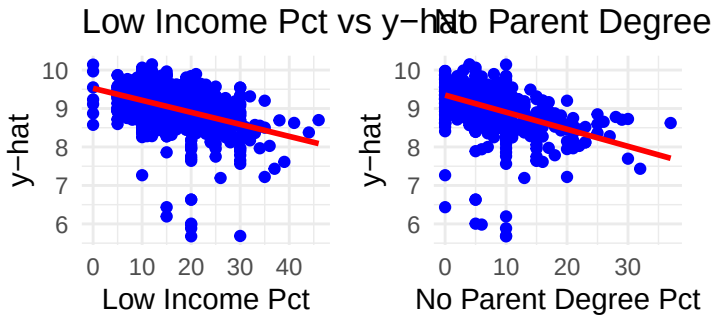
estimated parameters is a moderate to large fraction of the sample size. According to the rule of thumb provided by Burnham and Anderson (2004), AICc is preferred over AIC when the sample size $n \leq 40(p + 2)$ where p is the number of predictors. Based on this criterion, our dataset includes around eight hundred observations and only four predictors, which indicates that AIC is a more suitable measure for evaluating our model.

A Appendix

```
library(gridExtra)

# Add fitted values to df_clean
df_clean$fitted_values <- model$fitted.values

# Function for ggplot with regression line
ggplot_with_abline <- function(df, xcol, xlabel) {
  ggplot(df, aes(x = .data[[xcol]], y = fitted_values)) +
    geom_point(color = "blue") +
    geom_smooth(method = "lm", color = "red", se = FALSE) +
    labs(title = paste(xlabel, "vs y-hat"), x = xlabel, y = "y-hat") +
    theme_minimal()
}

# Create plots
 # Create boxplot for School_Type
p1 <- ggplot(df_clean, aes(x = School_Type, y = fitted_values)) +
  geom_boxplot(aes(fill = School_Type), outlier.colour = "red", outlier.shape = 16, outlier.s
  labs(title = "School Type vs y-hat", x = "School Type", y = "y-hat") +
  theme_minimal()

# Create boxplot for School_Language
p2 <- ggplot(df_clean, aes(x = School_Language, y = fitted_values)) +
  geom_boxplot(aes(fill = School_Language), outlier.colour = "red", outlier.shape = 16, outli
  labs(title = "School Language vs y-hat", x = "School Language", y = "y-hat") +
  theme_minimal()

p3 <- ggplot_with_abline(df_clean, "Special_Ed_Pct", "Special Ed Pct")
p4 <- ggplot_with_abline(df_clean, "Low_Income_Pct", "Low Income Pct")


p5 <- ggplot_with_abline(df_clean, "No_Parent_Degree_Pct", "No Parent Degree Pct")

# Arrange plots in a grid
grid.arrange(p1, p2, p3, p4, p5, ncol = 3, nrow = 2)
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```


School Type vs y−hat
School Language vs y−hat
Special Ed Pct vs
y−hat
School_Type
Catholic
Public
School_Language
English
French
School Type
School Language
Special Ed Pct


Low Income Pct vs y−hat
No Parent Degree Pct vs y−hat
y−hat
Low Income Pct
No Parent Degree Pct

References

Aseery, Ahmad. 2024. "Enhancing Learners' Motivation and Engagement in Religious Education Classes at Elementary Levels." *British Journal of Religious Education* 46 (1): 43–58.

Bernhofer, Juliana, and Mirco Tonin. 2022. "The Effect of the Language of Instruction on Academic Performance." *Labour Economics* 78: 102218.

Burnham, Kenneth P, and David R Anderson. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods & Research* 33 (2): 261–304.

Cheema, Jehanzeb Rashid. 2024. "Difference in Literacy Between Private and Public Schools: Evidence from a Survey of 61 Economies." *International Journal of Research in Education and Science* 10 (2): 218–40.

Davis-Kean, Pamela E, Lauren A Tighe, and Nicholas E Waters. 2021. "The Role of Parent Educational Attainment in Parenting and Children's Development." *Current Directions in Psychological Science* 30 (2): 186–92.

Nadeem, Tahir, Nasreen Akhtar, and Masood Ahmad. 2021. "A Study of the Relationship Between Family Income and Literacy Level." *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH* 3 (2): 59–69.

Ontario, Government of. 2024a. "Find Your School." https://www.ontario.ca/page/find-your-school.

———. 2024b. "School Information and Student Demographics." https://data.ontario.ca/dataset/school-information-and-student-demographics/resource/e0e90bd5-d662-401a-a6d2-60d69ac89d14.

Vaughn, Sharon, and Jeanne Wanzek. 2014. "Intensive Interventions in Reading for Students with Reading Disabilities: Meaningful Impacts." *Learning Disabilities Research & Practice* 29 (2): 46–53.

Zhang, Feng, Ying Jiang, Hua Ming, Yi Ren, Lei Wang, and Silin Huang. 2020. "Family Socio-Economic Status and Children's Academic Achievement: The Different Roles of Parental Academic Involvement and Subjective Social Mobility." *British Journal of Educational Psychology* 90 (3): 561–79.