Evaluation of EQAO Student Achievement Across Ontario Schools

Reshma Vangala, Gabriel Moffat, Hussein Eisa | Foundations of Data Science | October 2025

1. Abstract

This project analyzes Ontario's Board and School Information and EQAO student achievements for grades 3, 6, 9 and 10 (2023–2024). EQAO stands for the Education Quality and Accountability Office, an organization in Ontario that checks how well students are meeting provincial standards in reading, writing, and math. We discovered that the percentage of students gifted, percentage of students receiving special education, and socio-economic background are strong predictors. We also saw that schools that perform well in one subject generally perform well in others. Moreover, for Grade 3, a lower percentage of students met the standard for writing and math. And for Grade 6, a way lower percentage of students met the standard for just math. And a higher percentage of students met the standard for reading. And Grade 9 also shows a low percentage of students met the standard for math. So, there was major improvement in reading and writing between Grade 3 and 6 so that could be due to development of the kids or perhaps it indicates a higher quality of teachers in higher grades. Also, perhaps the math content got more difficult. Furthermore, we see that there is negative change in Grades 3 and 6 reading and writing over 3 years, but positive change in Grades 3, 6, and 9 math and Grade 10 OSSLT over 3 years, so the Ontario Ministry of Education might want to address that.

2. Introduction

The main issue we focus on is educational equity. Not all schools and students have the same needs, but resources like teachers, funding, and support programs are limited. This project is grounded in the practical use of data analytics for social good. By linking demographic and performance data, the analysis moves beyond raw numbers and helps uncover patterns that are often overlooked in traditional policy discussions to improve student success. The findings aim to support evidence-based decision-making within the Ministry of Education and local boards by showing which factors consistently predict gaps in achievement.

2.2 Data Source and Scope

The dataset used in this project comes from the Ontario Data Catalogue, specifically the file **new_sif_data_table_2023_24prelim_en_august2025**. This dataset is published by the Ontario Ministry of Education and provides detailed information on schools across the province.

The unit of analysis in this project is the school. Each row represents an individual publicly funded school in Ontario, with associated demographic and achievement data.

The dataset covers the **2023–2024 academic year (preliminary release)** and includes:

- **School information**: school and board number, name, and type; school special condition code; school level; school language; grade range; address info; contact info; enrollment.
- **Demographic percentages**: students whose first language is not English or French, students who are newcomers to Canada from non-English or non-French speaking countries, students receiving special education services or identified as gifted, students living in low-income households, and students whose parents have no degree, diploma, or certificate.
- **Student achievement indicators (EQAO results)**: % of Grade 3 and Grade 6 students meeting the provincial standard in Reading, Writing, and Mathematics; % of Grade 9 students meeting the provincial standard in Mathematics; % of Grade 10 students passing the Ontario Secondary School Literacy Test (OSSLT) on their first attempt; change in achievement over the past three years for Reading, Writing, Mathematics, and OSSLT.

Evaluation of EQAO Student Achievement Across Ontario Schools

Reshma Vangala, Gabriel Moffat, Hussein Eisa | Foundations of Data Science | October 2025

3. Data Acquisition and Cleaning/Preprocessing

3.1 Data Acquisition and Initial Load

The dataset used in this project was obtained from the Ontario Data Catalogue. Specifically, the file **new_sif_data_table_2023_24prelim_en_august2025.xlsx** was downloaded in Microsoft Excel format.

For the analysis, the Python programming language was used, along with libraries that support data analysis and visualization. The following key libraries were imported: pandas, numpy, matplotlib.pyplot, seaborn.

The dataset was loaded into a **Pandas DataFrame** for inspection and further analysis using pd.read_excel.

3.2 Data Cleaning and Preprocessing:

Step 1: Renaming Columns

- The original dataset contains column names that are quite long and with spaces and special characters (e.g., *"School Level"*).
- To make analysis easier and avoid syntax issues in Python, spaces were replaced with underscores, and names were shortened, using df.rename(columns={},inplace=True)

Step 2: Checking Unique Values in Each Column

- To understand our data better, we looked at all the unique values in each column. This helps us find:
    - Typos or unexpected entries – for example, "Ont" instead of "Ontario"
    - Missing or placeholder values – for example, NaN, "NA", "SP", "N/R" or "N/D".
    - Incorrect data types – sometimes numbers are stored as text.

  By checking unique values, we can decide which data needs to be cleaned or standardized before analysis.
- ➔ We applied df[col].unique(), with format, using a for loop through the columns in Python.

Step 3: Handling Special Values in the Enrolment Column

While checking unique values, we noticed that the **Enrolment** column contains the value "SP". According to the Ontario school glossary, "SP" represents small populations, typically **between 1 and 49 students**.

To handle this, we replaced "SP" with a random number between 1 and 49 so that the column remains numeric and usable for analysis. See Appendix A for Python code used.

Purpose of this step:

- Convert special "SP" values into numeric form.
- Maintain realistic enrolment numbers while making the data analysis-ready.

Step 4: Removing Rows with Missing or Special Values

Next, we focused on columns 24–47, which contain information that was sometimes not reported or not available. These columns may have values like "SP", "NA", "N/R", "N/D", or Null. According to the Ontario school glossary, these values indicate data not provided or suppressed for privacy or other reasons. Since this information cannot be used for analysis, we decided to remove rows that contained these values for all the achievement columns. See Appendix B for the Python code applied.

Purpose of this step:

- Remove rows with missing or suppressed data that cannot be analyzed.
- Ensure the dataset contains only reliable and usable information.

Dataset Shape After Cleaning Missing and Suppressed Values

After removing all rows that contained missing or suppressed values ("SP", "NA", "N/R", "N/D", or Null) from the relevant columns, the dataset now has 4,707 rows and 49 columns.

Step 5: Converting Percentage Columns to Decimals

We noticed that many percentage columns still contained values as strings (like "75%") or numbers greater than 1 (e.g., 75 instead of 0.75). To standardize these columns for analysis, we converted all percentages to decimal format. See Appendix C for the function used.

Step 6: Checking Data Types Before Conversion

Before converting columns to the appropriate data types, we first inspected the dataset using df.info() to see the current types. This helps identify columns that should be converted to numeric or categorical for analysis.

Step 7: Converting Columns to Appropriate Data Types

To optimize memory usage and improve performance, we converted certain columns to more suitable data types.

- Categorical Conversion – Columns with repeated text values, such as school type, board type, province, municipality, and city, were converted to the category data type. This reduces memory usage and improves performance when analyzing repeated text values.
- Numeric Conversion – Columns containing numeric values, such as Latitude, Longitude, and Enrolment, were converted to numeric types. This ensures calculations, mapping, and analysis can be performed correctly.

Step 8: Converting Percentage Columns to Numeric Where Possible

Some columns, particularly percentage columns, still contain special values such as "NA", “SP”, “N/D”, "N/R". To prepare the data for analysis, we converted these columns to numeric where possible, while keeping the special values unchanged using errors=’ignore’ in pd.to_numeric. This allows calculations, summaries, and visualizations on numeric values, without altering placeholders for missing or suppressed data, which can provide a meaningful and more detailed explanation for anyone inspecting the data, as each designation has its reasoning as indicated in the Ontario school glossary.

Evaluation of EQAO Student Achievement Across Ontario Schools

Reshma Vangala, Gabriel Moffat, Hussein Eisa | Foundations of Data Science | October 2025

Purpose of this step:

- Ensure numeric values are properly formatted for analysis.
- Preserve "NA", "N/R", “N/D”, and “SP” to maintain data integrity.

Step 9: Checking Missing Values

We checked the total number of missing values (NaN) in each column using df.isnull().sum() to see where the data is incomplete and may need further handling.

Step 10: Handling Missing Contact Information

Some contact columns had missing values:

- **Building_Suite** (4,908) and **PO_Box** (4,831) – Dropped with df.drop because they have excessive missing values that are not essential for analysis. They really are just providing additional address info that will not be needed.
- **Phone_Number** (163) – Since the dataset covers only Ontario publicly funded schools,we filled missing phone numbers, using fillna, with the Ontario Ministry of Education's general contact number: 416-325-2929, assuming that callers will be redirected appropriately.
- **Fax_Number** (356) – We are adopting a similar approach as above and filling the missing fax numbers with the general fax number of the Ontario Ministry of Education: 416-325-2929. Although fax communication is more specialized - often used to send official or sensitive documents, we are making a reasonable assumption that the Ministry's main office can redirect incoming faxes to the appropriate school or department if needed.
- **School_Website** (101) – Filled missing URLs with the Ontario “Find Your School” page.

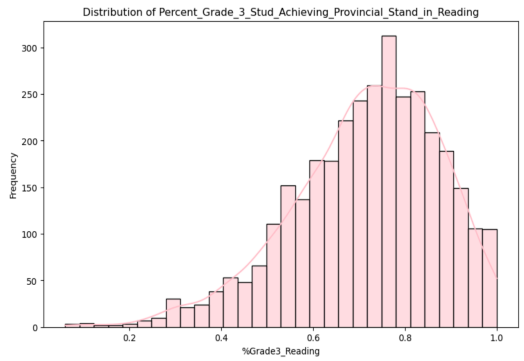
The above ensures all schools have contact info while keeping the dataset relevant and complete.

Step 12: Filling Missing EQAO Achievement Values

We did imputation: filled missing values for schools that include the grade in their Grade_Range based on the data distribution using sns.histplot:

- Right-skewed or left-skewed distributions – used the median to fill missing values.
- Approximately normal distributions – used the mean to fill missing values.

View Grade 3 reading sample output below. And the code used is in Appendix D.



Purpose of this step:

- Replace missing values with a representative statistic (median or mean) without affecting other data.
- Ensure all EQAO achievement value columns have complete data for analysis.

Step 13: Final Check for Missing Values

After filling in the applicable null values, the remaining missing values are only for cases where a school's grade range does not apply to that column (for example, Grade 3 results for high schools). These do not need to be filled, as they reflect the structure of the dataset.

At this point, data cleaning is complete, and we are ready to move on to analysis.

4. Analysis

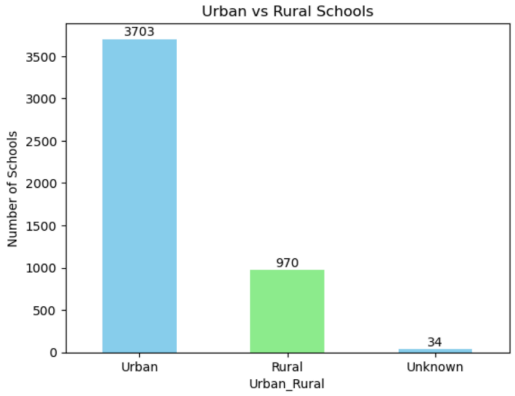
We posed the question: what category of students seem to be most likely to perform better on EQAO assessments?

4.1 Rural vs. Urban Performance

4.1.1 Rural vs. Urban Classification

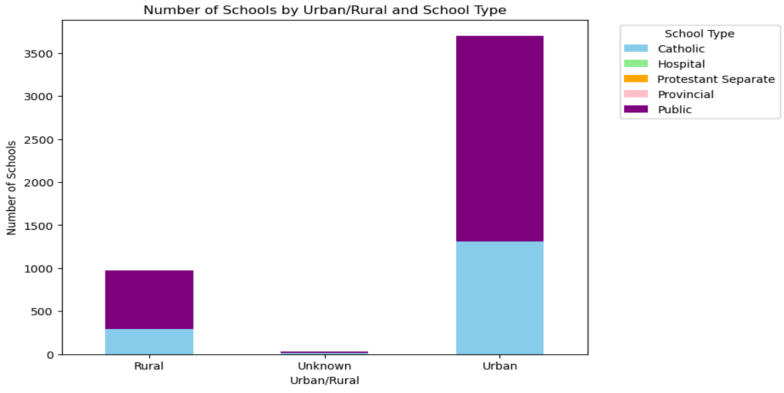
We analyzed student performance between schools located in rural and urban areas. To classify schools, we created a function and used the **Municipality** column and feature engineered a column **Urban_Rural**:

- Schools with "City of" or "Town of" in the name were classified as Urban.
- Schools with "Township," "Municipality," or similar terms were classified as Rural.



- Most schools are Urban.

4.1.2 School Distribution by Urban vs. Rural and School Type

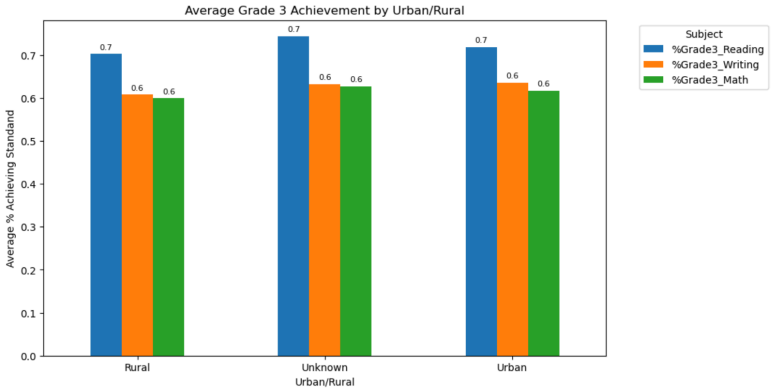


- Public and Catholic schools dominate both urban and rural areas, with public schools being more.

4.1.3 EQAO Assessment Analysis – Urban vs. Rural

We examined student performance across **Grades 3, 6, 9, and 10**, comparing **Urban and Rural schools** in **Reading, Writing, Math, and OSSLT outcomes**. We grouped by 'Urban_Rural' and applied the mean to the appropriate columns and plotted a bar graph. View sample output below:

Urban_Rural	%Grade3_Reading	%Grade3_Writing	%Grade3_Math
Rural	0.703035	0.607706	0.599358
Unknown	0.744231	0.632308	0.627200
Urban	0.719105	0.635981	0.616954



Summary Findings:

- For all grades, urban and rural schools seem to have similar performance, just that urban schools are slightly better.

Key Insight: Urban schools consistently perform slightly better than Rural schools, but the differences are small. This indicates that **location is not a strong factor** affecting student achievement.

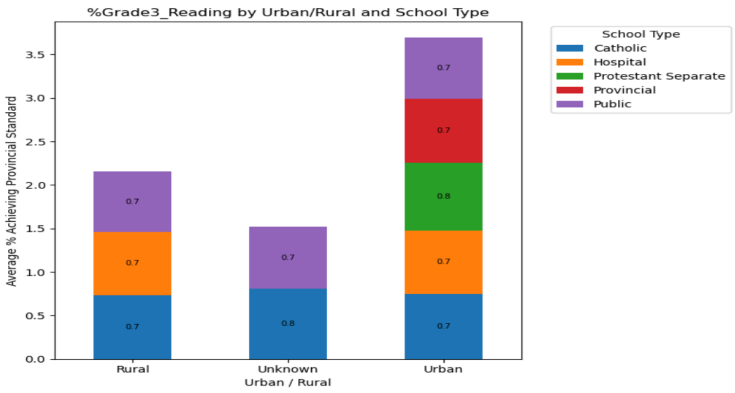
4.1.4 EQAO Assessment Analysis – Urban/Rural and School Type

We analyzed school performance across **Grades 3, 6, 9, and 10**, comparing **Urban vs. Rural schools** and different **school types** in **Reading, Writing, Math, and Grade 10 OSSLT results**. The code is similar to above with additional groupby by School_Type and plotting including stacked=true. The images below show sample table and graph used to draw conclusions:

Urban_Rural	School_Type	%Grade3_Reading	%Grade3_Writing	%Grade3_Math
Rural	Catholic	0.731682	0.629500	0.636273
	Hospital	0.730000	0.640000	0.620000
	Protestant Separate	NaN	NaN	NaN
	Provincial	NaN	NaN	NaN
	Public	0.690223	0.597935	0.582944
Unknown	Catholic	0.805556	0.714444	0.711250
	Hospital	NaN	NaN	NaN
	Protestant Separate	NaN	NaN	NaN
	Provincial	NaN	NaN	NaN
	Public	0.711765	0.588824	0.587647
Urban	Catholic	0.747009	0.659476	0.621464
	Hospital	0.730000	0.640000	0.620000
	Protestant Separate	0.780000	0.800000	0.710000
	Provincial	0.730000	0.640000	0.620000
	Public	0.703333	0.622630	0.614377

Evaluation of EQAO Student Achievement Across Ontario Schools

Reshma Vangala, Gabriel Moffat, Hussein Eisa | Foundations of Data Science | October 2025



Summary Findings:

- For all grades, Catholic schools perform better than Public schools in EQAO assessment, although only slightly so again.

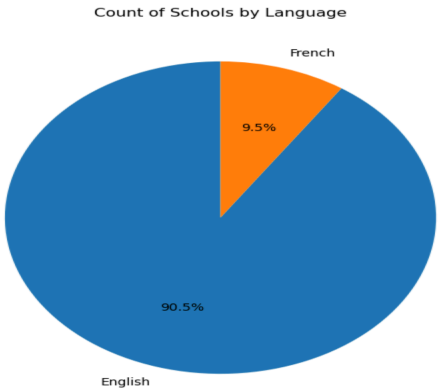
Key Insight: Catholic schools consistently perform slightly better than Public schools, but the differences are small. This indicates that **location and school type are not strong factors** affecting student achievement.

Other Findings from this Section:

- For Grade 3, a lower percentage of students met the standard for writing and math. And for Grade 6, a way lower percentage of students met the standard for just math. Moreover, a higher percentage of students met the standard for reading. And Grade 9 also shows a low percentage of students met the standard for math. So, there was major improvement in reading and writing between Grade 3 and 6 so that could be due to development of the kids or perhaps it indicates a higher quality of teachers in higher grades. And perhaps the math content got more difficult.
- A large percentage of students passed the OSSLT literacy test in their first attempt when they were in grade 10.

4.2 Language Background Performance

4.2.1 English vs. French School Classification



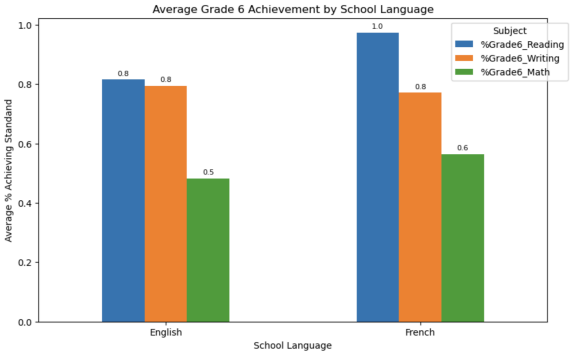
School Language Distribution:

- Most schools are English-language.
- French-language schools are few.

4.2.2 EQAO Assessment Analysis – English School vs. French School

We analyzed student achievement across **Grades 3, 6, 9, and 10** in **Reading, Writing, Math, and Grade 10 OSSLT results**, comparing **English- and French-language schools.** We grouped by 'School_Lang' and applied the mean to the appropriate columns and plotted a bar graph. View sample output:

School_Lang	%Grade6_Reading	%Grade6_Writing	%Grade6_Math
English	0.814953	0.794685	0.482051
French	0.973785	0.770956	0.564614



Summary Findings:

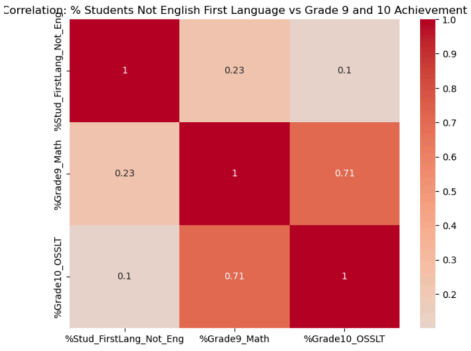
- For all grades, French schools perform better than English schools, although only slightly so, except for Grade 6 reading where it was more significant.

Key Insight: Again, like before, there is an indication that **language of school is not a strong factor** affecting student achievement.

4.2.3 EQAO Assessment Analysis – School Language and Students' First Language (English-language schools)

We examined whether the percentage of **students whose first language is not English** affects **performance** across **Grades 3, 6, 9, and 10**. See Appendix E for the sample code and view sample output:

	%Stud_FirstLang_Not_Eng	%Grade6_Reading	%Grade6_Writing	%Grade6_Math
%Stud_FirstLang_Not_Eng	1.000000	0.009167	0.079374	0.073577
%Grade6_Reading	0.009167	1.000000	0.827879	0.718643
%Grade6_Writing	0.079374	0.827879	1.000000	0.725634
%Grade6_Math	0.073577	0.718643	0.725634	1.000000



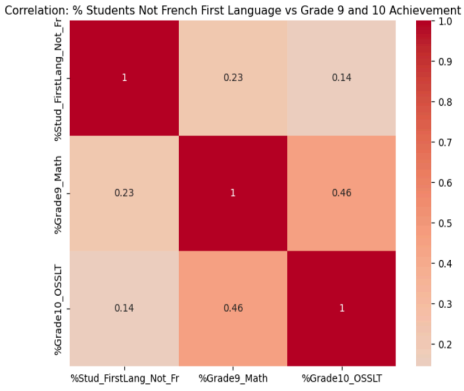
Summary Findings:

- There is a very weak correlation between Percentage of Students whose First Language is not English and Grade 3 and 6 achievement values. Although with grade 6 achievement values, the correlation between the writing and math and the Percentage of Students whose First Language is not English is similar and markedly stronger than the correlation between reading and the Percentage of Students whose First Language is not English.
- There is a weak positive correlation between Percentage of Students whose First Language is not English and Grade 9 and 10 achievement values, although more significant, so we can say that there is more of an impact in higher grades.

Key Insight: There seems to be an indication that **students' first language not being English in an English school is more of a factor** affecting student achievement **in higher grades**.

4.2.4 EQAO Assessment Analysis – School Language and Students' First Language (French-language schools)

We examined whether the percentage of **students whose first language is not French** affects **performance** across **Grades 3, 6, 9, and 10**. Code is similar to above, but modified for French schools. View sample output:



Summary Findings:

- There is a very weak correlation between Percentage of Students whose First Language is not French and Grade 3 and 6 achievement values, with the exception of Grade 6 math. Moreover, with grade 3, the correlation between the writing and math achievement values and the Percentage of Students whose First Language is not French is similar and markedly stronger than the correlation between the reading achievement value and the Percentage of Students whose First Language is not French.
- There is a weak positive, but more significant, correlation between Percentage of Students whose First Language is not French and Grade 9 and 10 achievement values, so we can say that there is more of an impact in higher grades.

Key Insight: There seems to be an indication that **students' first language not being French in a French school is more of a factor** affecting student achievement **in higher grades**, similar to the English schools findings.

4.2.5 EQAO Assessment Analysis – School Language and Students New to Canada from Non-English Countries in English Schools

To examine whether the percentage of students **new to Canada from non-English-speaking countries in English schools** affects **performance** in **Grades 3, 6, 9, and 10**. The code is similar to **4.2.3** but we used 'Percent_Stud_New_to_Canada_from_Non-English_Country' instead of 'Percent_Stud_First_Lang_Not_English'. View sample output:

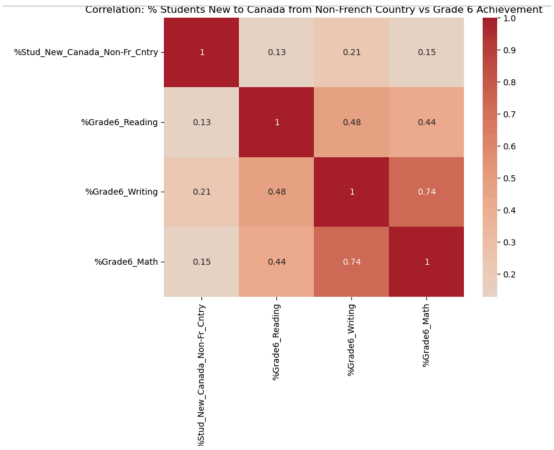
Summary Findings:

- There is a very weak negative correlation between Percentage of Students New to Canada from Non-English Countries and Grade 3, 6, 9 and 10 achievement values, with the exception of Grade 6 writing being very weak positive.

Key Insight: There seems to be an indication that **students' being new to Canada from non-English countries in English schools is not a strong factor** affecting student achievement.

4.2.6 EQAO Assessment Analysis – School Language and Students New to Canada from Non-French Countries in French Schools

To examine whether the percentage of students **new to Canada from non-French-speaking countries** affects **performance** in **Grades 3, 6, 9, and 10**. The code is similar to **4.2.4** but we used 'Percent_Stud_New_to_Canada_from_Non-French_Country' instead of 'Percent_Stud_First_Lang_Not_French'. View sample output:



Summary Findings:

- There is weak positive correlation between the Percentage of Students New to Canada from Non-French Country and Grade 3 and 6 achievement values, but there is very weak negative correlation between the Percentage of Students New to Canada from Non-French Country and Grade 9 and 10 achievement values. So it looks like there is more of an effect in French schools for younger grades.

Key Insight: There seems to be an indication that **students' being new to Canada from non-French countries in French schools is more of a factor** affecting student achievement **in lower grades**.

Other Findings from this Section:

- There is moderate to strong correlation between the achievement values themselves.

4.3 Special Education vs Non-Special Education Performance

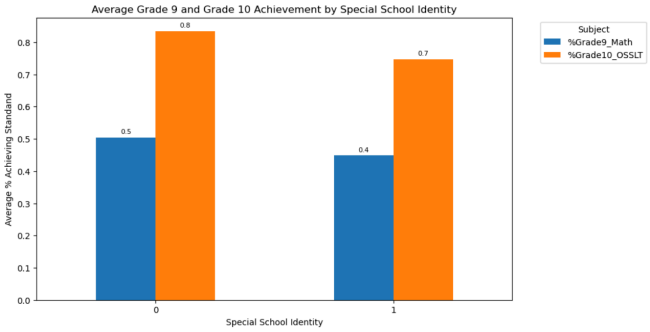
4.3.1 Special Education vs Non-Special Education Classification

We analyzed whether student performance differs between schools classified as special education and those that are not. To do this, we created a function and feature engineered a new column called **Spec_School_Identity** using the column **School_Spec_Cond_Code**.

- 0 = Regular school (School_Spec_Cond_Code is 'Not applicable'; 4629)
- 1 = Special education school (78)

4.3.2 EQAO Assessment Analysis – Special School Identity

We examined student performance across **Grades 3, 6, 9, and 10**, comparing **Special and Non-Special schools** in **Reading, Writing, Math, and OSSLT outcomes**. Code is similar to **4.1.3**, except that we grouped by 'Spec_School_Identity'. View sample output:



Summary Findings:

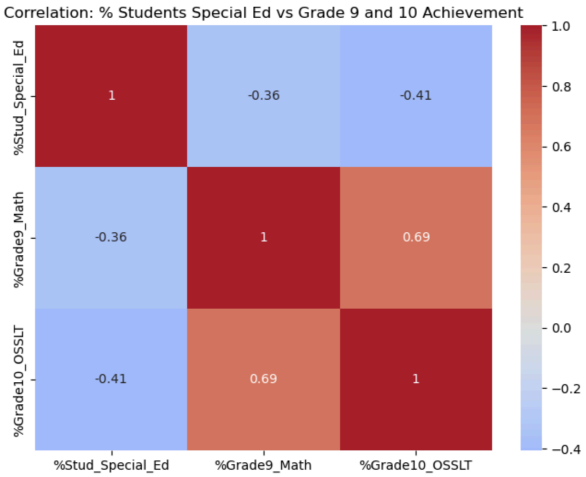
- For grades 3 and 6, Special and non-Special schools show very similar performance. However, special schools perform slightly higher, which is not expected.

- For grades 9 and 10, it is the reverse, Special and non-Special schools show very similar performance, but non special schools perform slightly higher.

Key Insight: Special and non special school differentiation has relatively minimal impact on performance, although we can note that special schools perform slightly better in lower grades while non special schools perform slightly better in higher grades. This indicates that **special school identification is not a strong factor** affecting student achievement.

4.3.3 EQAO Assessment Analysis – Percent Receiving Special Education

We examined the relationship between the **percentage of students receiving Special Education** and school performance in **Reading, Writing, Math, and OSSLT** across **Grades 3, 6, 9, and 10**. Code is similar to **4.2.3** except that we used 'Percent_Stud_Receiving_Special_Ed' instead of 'Percent_Stud_First_Lang_Not_English', and there was no differentiation between English and French schools. View sample output:



Summary Findings:

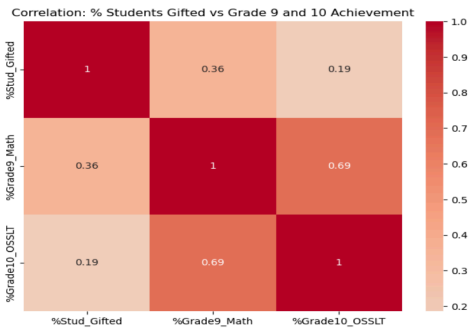
- There is a negative correlation between Percentage of Students who receive special ed and performance, starting with being weak to becoming moderate as the grades increase.
- So we can see that classification based on Spec_School_Identity was not sufficient.

Key Insight: Percentage of students receiving special education has somewhat of an impact on performance. This indicates that **the percentage of students receiving special education is a stronger factor** affecting student achievement.

4.4 Gifted Students Performance

4.4.1 EQAO Assessment Analysis – Gifted Students

We examined the relationship between the **percentage of students who are gifted** and school performance in **Reading, Writing, Math, and OSSLT** across **Grades 3, 6, 9, and 10**. Code is similar to **4.2.3** except that we used 'Percent_Stud_Gifted' instead of 'Percent_Stud_First_Lang_Not_English', and there was no differentiation between English and French schools. View sample output:



Summary Findings:

- There is a positive correlation between Percentage of Students who are gifted and performance, starting with being very weak to becoming weak and moderate as the grades increase.

Key Insight: Percentage of students who are gifted has more of an impact on higher grades performance. This indicates that **student giftedness is a stronger factor** affecting student achievement in higher grades.

4.5 Student Enrolment Performance

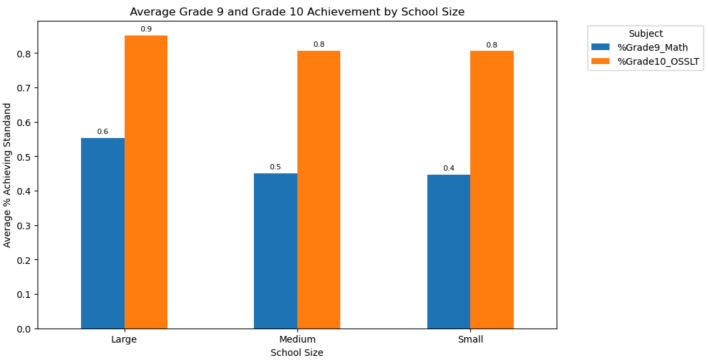
4.5.1 School Size Classification

We created a function and feature engineered a new column, **School_Size**, using the column **enrolment,** doing binning:

- 0 - 400 = Small (2697)
- 401 - 800 = Medium (1539)
- 801 - 2695 = Large (471)

4.5.2 EQAO Assessment Analysis – School Size

We examined the relationship between the **school size** and school performance in **Reading, Writing, Math, and OSSLT** across **Grades 3, 6, 9, and 10**. Code is similar to **4.1.3**, except that we grouped by 'School_Size'. View sample output:



Summary Findings:

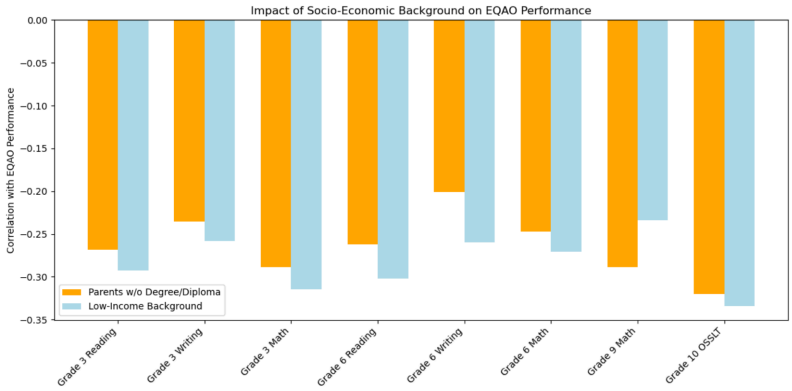
- Regardless of size, schools seem to have similar performance, but it is notable that Grade, 6, 9, and 10 reflect slight increase in performance going from Small, Medium, to Large schools, which is contrary to the expectation that smaller class ratios would perform better.

Key Insight: School Size has more of an impact on Grades 6, 9, and 10 performance. This indicates that **school size is a bit of a factor** affecting student achievement outside of Grade 3 years, with improvement from Small, to Medium, to Large school size.

4.6 Socio-Economic Background Performance

4.6.1 EQAO Assessment Analysis – Low-Income Background and Parent Education

We examined the relationship between the **percentage of school-aged children living in low-income households**, the **percentage of students whose parents have no degree, diploma, or certificate**, and school performance in **Reading, Writing, Math, and OSSLT** across **Grades 3, 6, 9, and 10**. We obtained correlation in the same manner that we have been doing, and plotted a bar graph with all achievement values at once. View output:



Summary Findings:

- Across all grades, the correlations are weak and negative, meaning that as the percentage of low-income students and parents without higher education increases, the percentage of students meeting provincial standards slightly decreases, but note that the correlations are more substantial than earlier ones that we have seen.

Key Insight: This indicates that **socio-economic background is perhaps the most significant factor** affecting student achievement, with parental obtainment of higher education playing more of a role.

4.7 Low-Income Background Performance

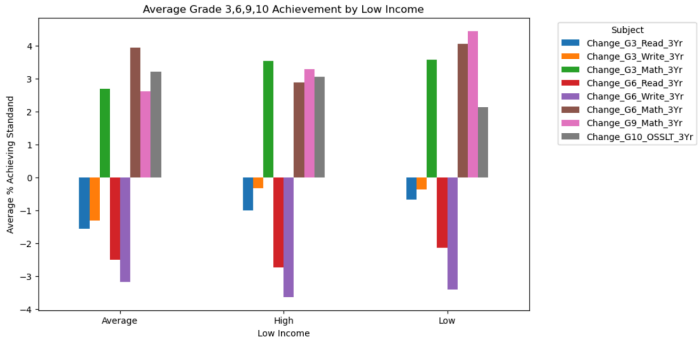
4.7.1 Low Income Number Classification

We created a function and feature engineered a new column, **Low_Income_Num**, using **Percent_School-Aged_Children_Live_in_Low-Income_Households,** doing binning**:**

- <= 0.05 = Low (1600)
- <= 0.1 = Average (1955)
- > 0.1 = High (1152)

4.7.2 EQAO Assessment Analysis – Low Income Number

We examined the relationship between the **low income number** and change in school performance over three years in **Reading, Writing, Math, and OSSLT** across **Grades 3, 6, 9, and 10**. Code is similar to what we have been doing. View output:



Summary Findings:

- Average low income numbers did worst in grade 3 change in achievement values. High low income numbers did worst in grade 6 change in achievement values. Average low income numbers had the lowest positive change in grade 9 math. And low low income numbers had the lowest positive change in grade 10 OSSLT.

Key Insight: This indicates that the Ontario Ministry of Education needs to focus on schools with a greater percentage of children living in low-income households from grades 3 to 9.

Other Findings From this Section:

- We see that there is negative change in Grades 3 and 6 reading and writing over 3 years, but positive change in Grades 3, 6, and 9 math and Grade 10 OSSLT over 3 years, so the Ontario Ministry of Education might want to address that.

5. Final Output: Key Factors Influencing EQAO Performance

I. Weak Predictors

- **Location (Urban vs. Rural):** Urban schools perform slightly better, but the difference is small.
- **School Type (Public vs. Catholic):** Catholic schools perform slightly better, but again the difference is small.
- **School Language (English vs. French):** French schools perform slightly better, but again the difference is small.

II. Moderate Predictors

- **First Language Not Matching School Language:** Slight effect in Grades 9 & 10 for both English and French schools.
- **Students New to Canada from Non-English and French Countries:** Negligible effect in English schools; slight impact in lower grades (3 & 6) for French schools.
- **School Size (Small, Medium, Large):** Regardless of size, schools seem to have similar performance, but it is notable that Grades 6, 9, and 10 reflect slight increase in performance going from Small, Medium, to Large schools, which is contrary to the expectation that smaller class ratios would perform better.

III. Strong Predictors

- **Percentage of Students Gifted**: There is a notable positive correlation in Grade 9 math.
- **Percentage of Students Receiving Special Education:** The negative correlation becomes more significant as the grades increase.
- **Socio-Economic Background:** This is where we saw the strongest negative correlations, with parents having not obtained higher education playing more of a role.

IV. Overall Trends

- For Grade 3 achievement values, a lower percentage of students met the standard for writing and math. And for Grade 6 achievement values, a way lower percentage of students met the standard for just math. Moreover, a higher percentage of students met the standard for reading. And the Grade 9 achievement value also shows a low percentage of students met the standard for math. So, there was major improvement in reading and writing between Grade 3 and 6 so that could be due to development of the kids or perhaps it indicates a higher quality of teachers in higher grades. And perhaps the math content got more difficult.
- We see that there is negative change in Grades 3 and 6 reading and writing over 3 years, but positive change in Grades 3, 6, and 9 math and Grade 10 OSSLT over 3 years, so the Ontario Ministry of Education might want to address that.
- Schools that perform well in one subject generally perform well in others.

6. Special Note on Columns Not Used for Analysis

We didn't use Board Number and Board Name because there are a lot of unique values, and we wouldn't be able to look at graphs using them properly; we didn't use Board Type because it is similar to School Type, which we used; we didn't use School Number and School Name because it is just a form of identifying schools; we didn't use school level because we used grade range, and grade range is more specific; we didn't use Street, City and Postal Code because we used Municipality and that was sufficient in doing analysis based on location. Also Province is all Ontario so it doesn't tell much for analysis; we didn't use Phone Number, Fax Number, School Website, and Board Website because it is just contact information; we considered using Latitude and Longitude, but ended up not doing so because we felt like we had done sufficient analysis based on location.

7. Appendix

A.

```
#replacing SP in Enrolment with values between 1 and 49 based on glossary (https://www.app.edu.gov.on.ca/eng/sift/glossary.asp)
df['Enrolment'] = df['Enrolment'].apply(lambda x: np.random.randint(1, 50) if x == 'SP' else x)
```

B.

```
#getting rows where SP,NA,N/D,N/R,null values exist in achievement columns
selection1 = (df[cols].isin(['SP', 'NA', 'N/D', 'N/R']) | df[cols].isna()).all(axis=1)

bad_rows_df = df[selection1] #data frame
```

```
df = df[~selection1].copy() #removing rows above
```

C.

```
#function to change percentages to decimals
def percent_fix(val):
    if pd.isna(val):
        return np.nan
    if isinstance(val, str) and val.endswith('%'):
        return float(val.strip('%')) / 100
    if isinstance(val, str):
        try:
            num = float(val)
            if num > 1:
                return num / 100
            return num
        except ValueError:
            return val
    else:
        return val

#fixing columns
for col in cols_to_fix:
    df[col] = df[col].apply(percent_fix)
```

D.

```
#to fill null with median
df.loc[df['Grade_Range'].isin(['JK-8','JK-6','1-8','JK-3','JK-4','JK-5','K-8', 'K-6', '1-5', '1-6']), '%Grade3_Reading'] = (
    df.loc[df['Grade_Range'].isin(['JK-8','JK-6','1-8','JK-3','JK-4','JK-5','K-8', 'K-6', '1-5', '1-6']), '%Grade3_Reading']
        .fillna(plot_data1.median())
)
```

E.

```
english_schools = df[df['School_Lang'] == 'English'].copy()
performance_cols = [
    '%Grade3_Reading',
    '%Grade3_Writing',
    '%Grade3_Math'
]
english_schools[['%Stud_FirstLang_Not_Eng'] + performance_cols] = english_schools[['%Stud_FirstLang_Not_Eng'] + performance_cols].apply(
    pd.to_numeric, errors='coerce')
corr_matrix = english_schools[['%Stud_FirstLang_Not_Eng'] + performance_cols].corr()
corr_matrix
```

```
#plotting scatter
fig, axes = plt.subplots(1, 3, figsize=(18,5), sharex=True, sharey=False)

for ax, y_col in zip(axes, performance_cols):
    sns.scatterplot(
        data=english_schools,
        x='%Stud_FirstLang_Not_Eng',
        y=y_col,
        alpha=0.7,
        ax=ax
    )
    ax.set_title(f"{y_col.split('_')[-1]} vs % Non-English First Language")
    ax.set_xlabel("% Non-English First Language")
```

```
#plotting heat map
plt.figure(figsize=(8,6))
sns.heatmap(
    corr_matrix,
    annot=True,
    cmap='coolwarm',
    center=0
)
plt.title("Correlation: % Students Not English First Language vs Grade 3 Achievement")
```