Association of Factors Related to Elementary Student EQAO Outcomes

# Introduction

## Context and research question

The academic performance of a school is believed to influence the academic achievement of its students. Choosing a school with better academic performance is often seen as a significant means of contributing to a child’s long-term academic success.

School academic performance can be reviewed using yearly EQAO scores and socioeconomic factors gathered during the EQAO evaluation process, such as parental income and education.

Non-socioeconomic factors also play a role. Non-private Ontario schools are divided into Catholic and Public boards. Various French Immersion programs are often provided to schools in both boards. These additional factors can be considered when creating a predictive model of EQAO scores.

The main research question of this paper is to determine which factors, as gathered during the EQAO survey process, are meaningfully associated with the highest fifth of EQAO performance in reading, writing and mathematics for Ontario Elementary students, in both Grade 3 and 6.

## Solution Overview

The proposed approach will aggregate EQAO scores and related factors available at the school and board level into a single table. Factors will be correlated to EQAO performance at both the Grade 3 and Grade 6 level during the 2017/2018 school year using association rules. Meaningful factors will be used to predict EQAO performance.

To evaluate the research question, the following main steps are proposed:

* Identify and obtain datasets
* Remove redundant and out-of-scope columns from datasets
* Merge datasets
* Convert required columns to factors and treat outlier and unbalanced values
* Generate association rules to identify relationship of factors to the EQAO scores
* Build and test a model to predict EQAO score

# Literature Review

## Reviewed Papers

1. Student Cohort Tracking Studies: Impacts of Gender, SES and Early Achievement on Students’ EQAO Performance
   1. <http://training.compassforsuccess.ca/pluginfile.php/200/mod_resource/content/0/SKGrade3andGrade6CohortStudiesFinal.pdf>
   2. SKGrade3andGrade6CohortStudiesFinal.pdf
   3. Binary and multinomial logistic regression models were used to determine correlation between student achievement in Senior Kindergarten, Grade 3 and Grade 6 given their socio-economic status with their subsequent performance in EQAO tests. Considers how student performance from earlier years could show a delayed effect in EQAO test results.
2. Tree cover and species composition effects on academic performance of primary school students
   1. <https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/2007691121?pq-origsite=summon>
   2. out.pdf
   3. Analysis of the effect of greenspace on 387 elementary schools in Toronto on EQAO test results using general linear modeling. Adjusts for the effect of socioeconomic factors known to influence test outcomes such as parental income and education. Use of the tool ‘R’ to perform statistical analysis.

# Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: building explanations

1. <https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/228620424?pq-origsite=summon>
2. ProQuestDocuments-2019-10-08.pdf
3. Explores the factors affecting EQAO performance of Grade 6 French Immersion students. Specifically evaluates the effect of French Immersion on overall performance in Grade 3 and Grade 6 standardized testing. The impact of delayed performance over several years is explored from the perspective of ‘catching up’ to non-French Immersion students. Exploration of this delay effect is relevant to framing the scope of this paper.
4. A Knowledge Synthesis of Large-Scale Assessments Relative to Canadian Educators
5. <https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1075&context=major-papers>
6. A Knowledge Synthesis of Large-Scale Assessments Relative to Cana.pdf
7. Explores the relationship between educator (teacher and school) and EQAO scores to examine professional performance and school ranking. It looks to understand if EQAO performance can be used as a proxy to meaningfully interpret the capability of educators. This is relevant when considering gender ratio as an influence on EQAO scores.
8. Merging Large-Scale Assessment Data for Secondary Analysis: Experiences with EQAO’s Data
   1. <http://www.ccsenet.org/journal/index.php/jel/article/view/24260>
   2. 24260-87665-1-PB.pdf
   3. Examination of the process and effects of merging EQAO datasets using SQL including resolution of merge conflicts and incomplete data. Exploration of the methodology used to ensure data integrity and to handle missing data. Consideration of how merge challenges were managed pertain to this paper.
9. Parenting practices focusing on literacy: a study of cultural capital of kindergarten and first-grade students from low-income families
10. <https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/03004430/v189i0003/500_ppfolaafsflf.xml&sub=all>
11. 2019\_-\_Kathryn\_E\_Bojczyk\_-\_Parentingpracticesfocusingonliteracyastudyofcultur[retrieved\_2019-10-08].pdf
12. Examination of the statistical relationship between socioeconomic factors (“social capital”) and literacy in 198 urban American primary aged school children. Use of descriptive, correlation matrices and hierarchical regression analysis outline analytical techniques that were considered when developing this paper.
13. Emergent Metaliteracies: What the Xbox has to Offer the EQAO
14. <https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/08985898/v14i3-4/305_emwtxhtote.xml>
15. 2003\_-\_\_-\_EmergentMetaliteraciesWhattheXboxhastoOffertheEQAO[retrieved\_2019-10-08].pdf
16. Explores the potential limits of literacy results in EQAO within the broader effect of digital literacy (broadly defined to include pop media and gaming platforms). Suggests the need to consider the limits of models based on currently available EQAO factors. Exploration of these considerations is relevant to framing the context and scope of this paper.

# Dataset

## Data Source

* All data sourced from:
  + <https://www.ontario.ca/data/>
* All data is marked “Open” which makes it available for research purposes as well as general user per Ontario’s “Sharing Government Data” guidelines:
  + <https://www.ontario.ca/page/sharing-government-data>

## Overview of Data

* 6 Dependent Variables (highlighted in yellow)
* 37 Independent Variables
* Note: variables sorted by “Source”. See Source Detail below.
* Available on <https://github.com/ryerstan/CKME136>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column** | **Source** | **Variable** | **Type** | **Note** |
| Board Number | 1\_Copy | Index | Character |  |
| Board Type | 1\_Copy | Independent | Factor |  |
| School Number | 1\_Copy | Independent | Character |  |
| School Special Condition Code | 1\_Copy | Independent | Factor |  |
| School Language | 1\_Copy | Independent | Factor |  |
| Enrolment | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Whose First Language Is Not English | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Whose First Language Is Not French | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Who Are New to Canada from a Non-English Speaking Country | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Who Are New to Canada from a Non-French Speaking Country | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Receiving Special Education Services | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Identified as Gifted | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Grade 3 Students Achieving the Provincial Standard in Reading | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Percentage of Grade 3 Students Achieving the Provincial Standard in Writing | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Percentage of Grade 3 Students Achieving the Provincial Standard in Mathematics | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Percentage of Grade 6 Students Achieving the Provincial Standard in Reading | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Percentage of Grade 6 Students Achieving the Provincial Standard in Writing | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Percentage of Grade 6 Students Achieving the Provincial Standard in Mathematics | 1\_Copy | Dependent | Numeric | Will be converted to a factor |
| Change in Grade 3 Reading Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Change in Grade 3 Writing Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Change in Grade 3 Mathematics Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Change in Grade 6 Reading Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Change in Grade 6 Writing Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Change in Grade 6 Mathematics Achievement Over Three Years | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Children Who Live in Low-Income Households | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Percentage of Students Whose Parents Have Some University Education | 1\_Copy | Independent | Numeric | Will be converted to a factor |
| Junior Kindergarten Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Kindergarten Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 1 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 2 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 3 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 4 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 5 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 6 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 7 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Grade 8 Enrolment | 2\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Board Language | 3\_Copy | Independent | Factor |  |
| Elementary Male Educators | 3\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| Elementary Female Educators | 3\_Copy | Independent | Numeric | Will be converted to a factor as a relative measure to the board average |
| FSL Core | 4\_Copy | Independent | Factor | Binary |
| FSL Extended | 4\_Copy | Independent | Factor | Binary |
| FSL Immersion | 4\_Copy | Independent | Factor | Binary |

## Source Detail

### EQAO performance and socio-economic statistics

<https://www.ontario.ca/data/school-information-and-student-demographics>

**Original XLSX**

sif\_data\_table\_2017\_2018prelim\_en\_june

**Renamed XLSX**

1\_Source\_sif\_data\_table\_2017\_2018prelim\_en\_june

**CSV Copy**

1\_Copy

**Year Covered**

2017 to 2018

**Granularity**

Records at the School level

**# Records**

4912

**Description**

#Copy\_1 contains EQAO performance and socio-economic statistics

### Enrolment by grade for elementary schools by board

<https://www.ontario.ca/data/enrolment-grade-elementary-schools>

**Original XLSX**

enrolment\_by\_grade\_elementary\_schools\_en

**Renamed XLSX**

2\_Source\_enrolment\_by\_grade\_elementary\_schools\_en

**CSV Copy**

2\_Copy

**Year Covered**

2017 to 2018

**Granularity**

Records at the Board level

**# Records**

75

**Description**

#Copy\_5 contains enrolment by grade for elementary schools by board

### Teacher gender distribution by board

<https://www.ontario.ca/data/number-full-time-educators-gender-and-school-board-ontario>

**Original XLSX**

ftebyboard\_en\_1718

**Renamed XLSX**

3\_Source\_ftebyboard\_en\_1718

**CSV Copy**

3\_Copy

**Year Covered**

2017 to 2018

**Granularity**

Records at the Board level

**# Records**

76

**Description**

#Copy\_7 contains teacher gender distribution by board

### French as a Second Language Programming

<https://www.ontario.ca/data/french-second-language-enrolment>

**Original XLSX**

schools\_offering\_fsl\_en\_1516

**Renamed XLSX**

4\_Source\_schools\_offering\_fsl\_en\_1516

**CSV Copy**

4\_Copy

**Year Covered**

2015 to 2016

**Granularity**

Records at the School level

**# Records**

4269

**Description**

#Copy\_10 contains FSL programming (core, extended, immersion)

# Approach

## Block Diagram

## Block Diagram Details

### Step 1: Remove unnecessary columns

* Remove redundant and out-of-scope columns from datasets
* All columns not initially identified as a dependent or independent variable are removed
  + Examples include street address, school name and board website
  + Removed columns are referential only and are not defined as independent or dependent variables
  + Copy\_1, originally with 51 columns, 23 columns removed, 28 remaining
  + Copy\_2, originally with 13 columns, 2 columns removed, 11 remaining
  + Copy\_3, originally with 13 columns, 9 columns removed, 4 remaining
  + Copy\_4, originally with 13 columns, 8 columns removed, 5 remaining
  + Total of 42 columns removed
  + Total of 48 columns remaining before merge
* <https://github.com/ryerstan/CKME136>
  + See Merged.R

### Step 2: Merge Datasets

* Merge Datasets
* Copy\_1 and Copy\_4 were merged first via an inner join
  + Only rows where school IDs matched in both datasets were retained
  + Both of these datasets are at the school level of granularity
* The result of merging Copy\_1 and Copy\_4 were then merged via an inner join sequentially, first with Copy\_2 and then Copy\_3
  + Copy\_2 and Copy\_3 are at the board level of granularity
  + The resulting flat file (“Merged”) contains all merged data where only rows with school IDs and board IDs matched in all datasets were retained
  + Board information at the row level is replicated for each school that belongs to the same board
* Redundant School ID and Board ID columns were removed after each merge
  + At this stage, the merged dataset contains 44 columns and 3721 rows of data
* As a final step, the rows containing non-Elementary school data were removed as they are out-of-scope for the research question
  + After removing non-Elementary school data, the merged dataset contains 3064 rows
  + After removing the School.Level column (no longer required as it only contains the value “Elementary”), 43 columns remain (6 dependent and 37 independent variables)
* <https://github.com/ryerstan/CKME136>
  + See Merged.R
  + See Merged.csv

### Step 3: Column Factorization

* Convert required columns to factors and treat outlier and unbalanced values
* Remove Board.Type column: data available at school level under School.Type
* Remove School.Special.Condition column: only contained ‘Not Applicable”, except for 6 instances which referenced ‘Junior High School’ which is informational and can be ignored
* Remove School.Language column: only contains ‘English’
* Remove Board.Language column: only contains ‘English’
* For the following columns, NAs were imputed by calculating the Board mean, and if this was not available by overall mean
  + Percentage.of.Students.Whose.First.Language.Is.Not.English
    - 916 NA, 2147 numeric, 30% NA
  + Percentage.of.Students.Whose.First.Language.Is.Not.French
    - 5 NA, 3058 numeric, 0.16% NA
  + Percentage.of.Students.Who.Are.New.to.Canada.from.a.Non.English.Speaking.Country
    - 1607 NA, 1456 numeric, 52% NA
  + Percentage.of.Students.Who.Are.New.to.Canada.from.a.Non.French.Speaking.Country
    - 1603 NA, 1460 numeric, 52% NA
  + Percentage.of.Students.Receiving.Special.Education.Services
    - 57 NA, 3006 numeric, 1.9% NA
  + Percentage.of.Children.Who.Live.in.Low.Income.Households
    - 77 NA, 2986 numeric, 2.5% NA
  + Percentage.of.Students.Whose.Parents.Have.Some.University.Education
    - 51 NA, 3012 numeric, 1.7% NA
  + 7 NA, 3057 numeric, ~0% NA
* For the following columns, NAs were handled differently than replacement by means:
  + Percentage.of.Students.Identified.as.Gifted
    - 2859 NA, 204 numeric, 93% NA
    - Converted to binary
  + Elementary.Male.Educators and Elementary.Female.Educators
    - Replace NA with mean Board mean, and if this was not available by overall mean
      * Male
        + 2 NA, 3062 numeric, ~0% NA
      * Female
        + 0 NA, no imputation necassary
    - Replace both values at once because they are calculated as a ratio in new column Gender.Ratio.M.to.F
* All Grade enrolments were cleaned (comma removal)
  + The following new enrolment columns were created
  + Enrol\_sum\_all
    - Sum of all enrolments per grade per school
  + Enrol\_sum\_JKto3
    - Sum of all enrolments from JK to Grade 3 per school
  + Enrol\_sum\_4to6
    - Sum of all enrolments from Grade 4 to 6 per school
* All Grade 3 and Grade 6 EQAO rows for Reading, Writing and Mathematics (6 in total) are dependent variables and were removed if they contained NAs
  + Following this removal, 2390 rows remain
* Saved as Clean.csv and Clean.txt (Tab-Delimited)
  + <https://github.com/ryerstan/CKME136>
* Non-percentage numeric columns were scaled
  + - Enrol\_sum\_all
    - Enrol\_sum\_JKto3
    - Enrol\_sum\_4to6
  + All percentage numeric and scaled columns were factored using quantcut and the following scale:
    - Lowest, Low, Medium, High, Highest
* See Factored.csv and Factored.txt (Tab-Delimited)
  + <https://github.com/ryerstan/CKME136>

### Step 4: Generate Association Rules

* Factorized dataset ‘Final’ turned into transactions type dataset ‘trans\_Final’
* Summary of ‘trans\_Final’:
* labels
* 1 School.Type=Catholic
* 2 School.Type=Public
* 3 Enrolment=Lowest
* 4 Enrolment=Low
* 5 Enrolment=Medium
* 6 Enrolment=High
* 7 Enrolment=Highest
* 8 Percentage.of.Students.Whose.First.Language.Is.Not.English=Lowest
* 9 Percentage.of.Students.Whose.First.Language.Is.Not.English=Low
* 10 Percentage.of.Students.Whose.First.Language.Is.Not.English=Medium
* 11 Percentage.of.Students.Whose.First.Language.Is.Not.English=High
* 12 Percentage.of.Students.Whose.First.Language.Is.Not.English=Highest
* variables levels
* 1 School.Type Catholic
* 2 School.Type Public
* 3 Enrolment Lowest
* 4 Enrolment Low
* 5 Enrolment Medium
* 6 Enrolment High
* 7 Enrolment Highest
* 8 Percentage.of.Students.Whose.First.Language.Is.Not.English Lowest
* 9 Percentage.of.Students.Whose.First.Language.Is.Not.English Low
* 10 Percentage.of.Students.Whose.First.Language.Is.Not.English Medium
* 11 Percentage.of.Students.Whose.First.Language.Is.Not.English High
* 12 Percentage.of.Students.Whose.First.Language.Is.Not.English Highest
* > summary(trans\_Final)
* transactions as itemMatrix in sparse format with
* 2390 rows (elements/itemsets/transactions) and
* 100 columns (items) and a density of 0.23
* most frequent items:
* FSL.Core=YES
* 2284
* FSL.Extended=NO
* 2257
* Gifted=FALSE
* 2255
* Percentage.of.Students.Whose.First.Language.Is.Not.French=Low
* 1971
* FSL.Immersion=NO
* 1902
* (Other)
* 44301
* element (itemset/transaction) length distribution:
* sizes
* 23
* 2390
* Min. 1st Qu. Median Mean 3rd Qu. Max.
* 23 23 23 23 23 23
* includes extended item information - examples:
* labels variables levels
* 1 School.Type=Catholic School.Type Catholic
* 2 School.Type=Public School.Type Public
* 3 Enrolment=Lowest Enrolment Lowest
* includes extended transaction information - examples:
* transactionID
* 1 1
* 2 2
* 3 3
* Apriori will be used to generate association rules in

### Step 5: Build Model

* Build and test a model to predict EQAO score
* Not completed at this time

# Results

To be populated in Final Project Report.

# Conclusions

To be populated in Final Project Report.