Project Student Success

How household income and parent education levels, and school enrolment sizes effect student achievement in provincial standards and examinations.

Najma Osman

09 February 2021

1 Project Overview

1.1 Early Exploration Results

To start, I plotted two scatter plots and two tables, moving toward answering Section questions 3, 5 and 6 from Section 1.3 (Key Questions).

1.1.1 What did you find?

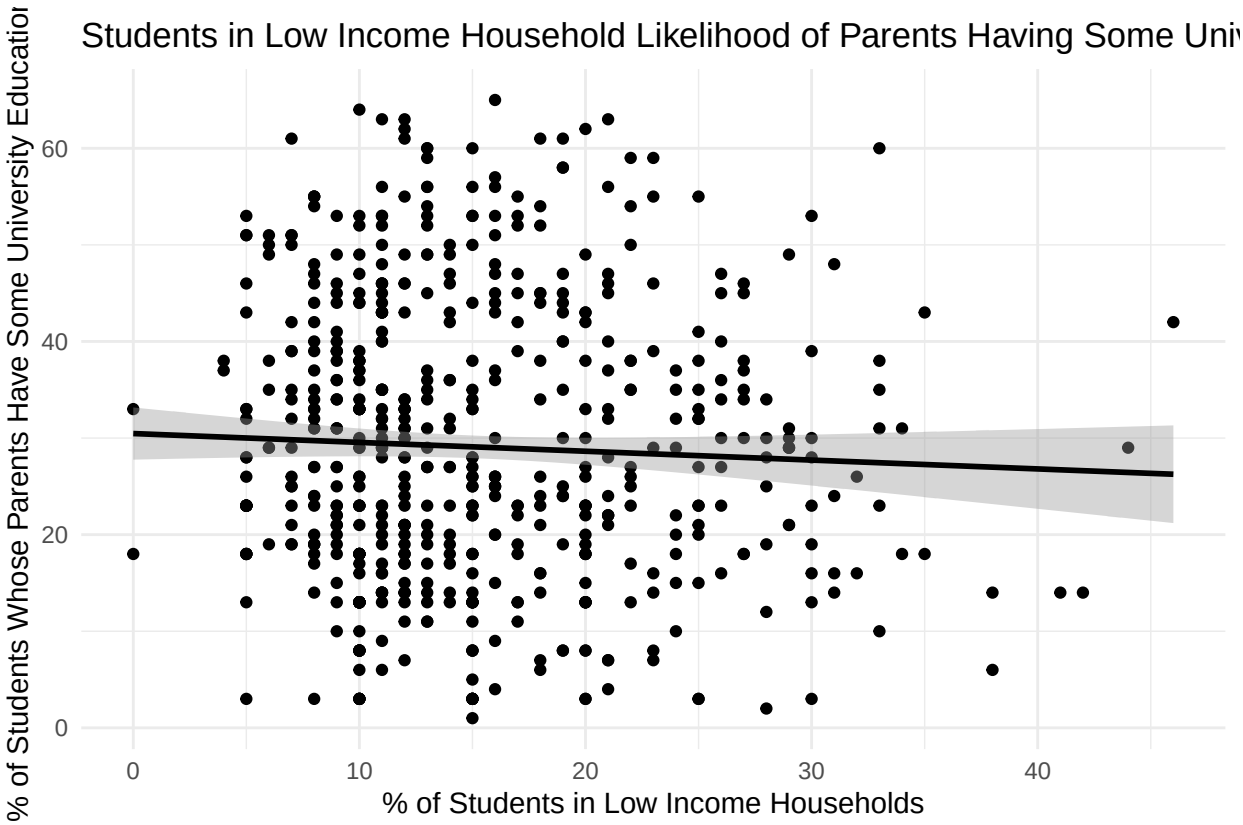

Figure 1

Figure 1 explored the relationship between the percentage of students in low income households and the percentage of students whose parents have some university education. By the best fit line, it seems incredibly small, but there is a slight downward trend in percentage of low income students when the percentage of students with parents who have some university education increases. However, the spread is far from uniform.

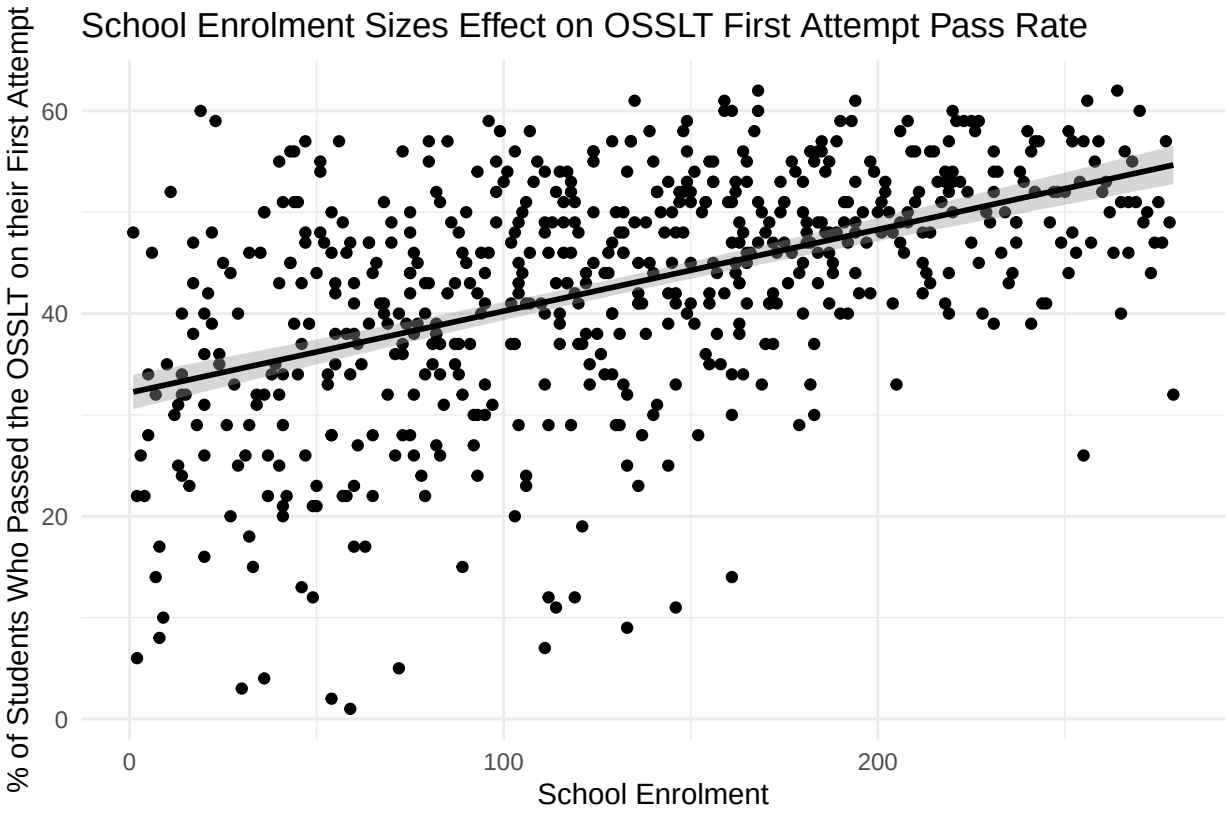

Figure 2

In comparison, Figure 2 appears to have a more noticeable trend, where we see higher student enrolment decreasing the percentage of students who passed the OSSLT on their first attempt. Again, the spread is large, and there's a significant amount of outliers, so maybe averaging schools out to the district school board level will present a more consistent picture.

Schools with Highest OSSLT Pass Rate on First Try
and their corresponding achievement of provincial math standards.

Board Name	School Name	osslt	academic_math	applied_math
York CDSB	Father Michael McGivney Catholic High School	97	88	39
York Region DSB	Pierre Elliott Trudeau High School	97	95	37
Ottawa-Carleton DSB	Lisgar Collegiate Institute	96	95	47
Toronto DSB	Agincourt Collegiate Institute	96	87	33
Toronto DSB	Etobicoke School of the Arts	96	84	55
York Region DSB	Bayview Secondary School	96	96	45
Toronto DSB	Earl Haig Secondary School	95	95	24
Toronto DSB	Malvern Collegiate Institute	95	86	39
Toronto DSB	Rosedale Heights School of the Arts	95	85	53
Upper Canada DSB	Vankleek Hill Collegiate Institute	95	79	67

Interestingly, the table (Table 1) that looks to the top 10 schools for the Grade 10 Literacy Test (OSSLT)

pass rates shows us a few things:

1. 8 out of 10 of the schools are located in the Greater Toronto Area (York CDSB, York Region DSB, and Toronto DSB),
2. These schools also have an exceptional amount of their students who are at or above the provincial standard for academic mathematics,
3. But they do not perform as well with applied mathematics.

Schools with Lowest OSSLT Pass Rate on First Try
and their corresponding achievement of provincial math standards.

Board Name	School Name	osslt	academic_math	applied_math
York Region DSB	Woodbridge College	0.83	0.73	0.29
Keewatin-Patricia DSB	Sioux North District High School	29.00	55.00	14.00
Thames Valley DSB	Arthur Voaden Secondary School	32.00	31.00	32.00
Durham DSB	G L Roberts Collegiate and Vocational Institute	38.00	88.00	59.00
Greater Essex County DSB	Westview Freedom Academy Secondary School	39.00	78.00	26.00
Superior-Greenstone DSB	Nipigon Red Rock District High School	40.00	61.00	67.00
Thames Valley DSB	Montcalm Secondary School	41.00	79.00	50.00
DSB Ontario North East	Cochrane High School	42.00	36.00	15.00
Thames Valley DSB	College Avenue Secondary School	43.00	64.00	32.00
Superior-Greenstone DSB	Geraldton Composite School	44.00	73.00	65.00

As for the bottom 10:

1. 8 out of 10 of the schools perform better in academic math than the OSSLT,
2. and only 2 out of 10 perform better in applied mathematics than academic mathematics and/or the OSSLT.

1.1.2 Next Steps

Next steps include aggregating/summarising the data at the city, or even the board level, so we can see a clearer picture of where the province may need to put more resources at a municipal or regional level by adjusting these tables or graphs to that.

1.2 Key Data Elements/Fields

- % of students whose parents have some university education
- % of school aged children who live in low income households
- % of students that passed the grade 10 OSSLT on their first attempt
- Change in grade 10 OSSLT literacy achievement over three years?
- % of grade 9 students achieving the provincial standard in academic mathematics
- Change in grade 9 academic mathematics achievement over three years?
- % of grade 9 students achieving the provincial standard in applied mathematics
- Change in grade 9 applied mathematics achievement over three years

1.3 Key Questions

1. Does the percentage of students who live in low income households affect student achievement in academic and applied mathematics for the worst? What about the inverse?
    a. What about percentage of students passing the OSSLT on the first try?
2. What schools have the most students reaching the provincial standard in mathematics?
    a. Do those schools also have the highest percentages of students passing the OSSLT? Or is there an inverse relationship between mathematics and literacy achievement? (e.g., increased mathematics achievement over 3 years but decreased instance of OSSLT being passed on first attempt)
3. What cities/school boards have the highest average of students passing the OSSLT on the first attempt?
4. Do schools with higher percentages of parents with some post secondary achievement have a higher instance of student giftedness, and a lower instance of students receiving special education services?
5. Does school enrolment size affect achievement in mathematics? Likelihood of passing the OSSLT on the first try?
6. Do schools with more low income students also have less parents who have had some post-secondary education?

2 Data

2.1 Source + Type

I've chosen a public dataset from Ontario's Data Catalogue. This dataset is on school information and student demographics in the province of Ontario, not including private schools, Education and Community Partnership Programs, or summer and night schools. It includes data on the following:

- board information
- school information
- grade 3 and 6 EQAO achievements for reading, writing and mathematics
- grade 9 EQAO academic and applied student achievements
- grade 10 OSSLT student achievement
- student demographic percentages on student parents, special education, first language spoken, and new students to Canada

The data is reported by:

- Board School Identification Database (BSID) and Ontario School Information System (OnSIS),
- OnSIS Preliminary 2018-2019 (Student Population)
- the Education Quality and Accountability Office (EQAO), specifically the 2018-2019 data,
- and the 2016 census.

2.2 Cleaning Requirements

Insofar, I've used the package janitor by Firke (2021) to modify the column names into snake case (variable_name) format, to access column names more easily, and reduced data using the tidyverse package by Wickham et al. (2019), which I'll go into detail about in section 2.2.6. Additional cleaning is required for converting columns into proper numeric values, rather than factors and/or characters as is the case for some. Click here to see a copy of the cleaned data.

2.2.1 Data Quality

The dataset is current (last updated December 2020, and generally updated on a monthly basis), has consistent units, is complete for all schools in the province of Ontario, and correct.

2.2.2 Anonymization

This dataset pulls from OnSIS and Statistics Canada, which both suppress results for variables based on school population size to protect student privacy. The following methods were used to ensure anonymity:

- randomly rounding percentages up or down depending on school enrolment,
  - 0 - NA
  - 1-49 - SP
  - 50-99 - round up or down to a multiple of 10
  - 100-499 - round up or down to a multiple of 5
  - 500-4,999 - round up or down to the ones digit
  - 5,000 + - round up or down to one decimal place
- not publicly reporting data where enrolment is less than 10.

2.2.3 Inconsistencies

- There are no name variations, or name changes (as of December 2020),
- no user-provided input, typos/spaces, and dates are consistent,
- and all columns utilising measurement (e.g., percentages) follow the same unit of measurement, and hold a single value,

2.2.4 Missing Data

There 4 instances of missing data in this set, which is already handled by the cataloguer:

- where student population information isn't available due to the school board not providing the data to the ministry (denoted by NA),
- where schools the school does not have EQAO results (denoted by ND),

- where the number of students participating is fewer than 10 and anonymity isn't ensured (denoted by NR),
- and where the results are repressed due to school enrolment of fewer than 50 students (denoted by SP).

Where school information is missing, such as fax numbers, it's been left blank and I've made no changes.

2.2.5 Outliers

All data is aggregated at the school level, none were omitted.

2.2.6 Unwanted Data

Since my research questions/interests largely focus on investigating the relationship of parent education levels and household income on student achievement in standardized examinations/provincial standards in Ontario's secondary school students, while factoring in gifted/special student population, and municipality, the following was done to reduce the dataset:

I removed rows:

- containing elementary schools,
- containing schools where the language is French, as they have different tests which are covered in a separate dataset,
- having codes NA, NR, or ND in columns to do with Grade 9 mathematics achievement of provincial standards and/or the Grade 10 literacy test (OSSLT),
- containing code SP in columns to do with household income and/or percentage of students whose parents have some university education.

I also removed the following columns from the dataset:

- location information: board number, board type, school number, building suite, P.O. Box, Street, postal code, municipality, province, latitude, longitude,
- school info: school type, school level, grade_range, enrolment, school special condition code (e.g., Alternative, Adult, NA), grade range,
- contact information: phone number, fax number, school website, board website
- testing information: grade 3 and 6 results (EQAO, achievement of provincial standard (percentage of student and change over 3 years))
- student info: percentage of students new to Canada from non English or non French speaking countries, percentage of students whose first language is not English or not French
- extract date

In the end, 14 columns remain of the original 51, and the data is reduced from 4887 records to 604 records, due to the removed rows mentioned above.

3 Expected Outcomes

With this data, and these questions, I expect I'll be able to tell a story on how the percentage of students in low income households, percentage of students whose parents have had some university education, and even school enrolment sizes affect school performance on achievement of provincial standards in Grade 9 mathematics, as well as the literacy test (OSSLT). That is, we can see how or if these dimensions result in higher or lower pass rates, or achievement of provincial standards, and then extrapolate on why this may be

the case (via research). For instance, if enrolment size appears to affect the school or board's achievement, we can potentially make a case for an issue of resources, or large class sizes. If the percentage of students in low income households, or the percentage of students whose parents have had some university education is higher for those schools with lower pass rate of the OSSLT, or lower instance of achievement of provincial standards in mathematics, we could potentially point to fewer resources at home. It's also possible to look to data at the board level, and see which boards perform exceptionally well or poorly, allowing us to see if the effects are only at a single school within the district or present throughout the district.

4 Expected Challenges

A major challenge with this analysis is that two of the primary variables being looked at are estimates: the percentage of students in low income households, and the percentage of students whose parents have some university education.

Both are calculated using student postal code data which is collected by the school, and then cross-referenced with the 2016 Census data from Statistics Canada. The measures are directly proportionate to the number of children from those different postal codes attending the school. For low income students, it's gathered by determining the percentage of school-aged children in households below the after-tax-low-income measure threshold (LIM-AT). Whereas for the university education metric, it's determined by seeing if one parent in the household (still within the postal code estimations) has a certificate, diploma, or degree from a university.

In addition to this, funding may explain some issues at the school and/or board level, but the allocation of funding from the province, to the boards (or from the boards), and then to the school doesn't appear to be transparent. For instance, access to the 2018-2019 budget plan for the Toronto District School Board is hidden behind an employee login page for the TDSB. Funding has to exist as a confounding variable at this point in time, since this dataset doesn't have any funding related columns.

5 Next Steps

My skills are largely technical since I have some working background in programming. I'm comfortable working either in R or Python to both clean and analyze datasets, and even come to conclusions or observations based on that initial analysis. However, while I can make the graphs and tables to be able to do that, I've never had a phenomenal eye when it comes to actually visualizing the data. Ideally, my team members will be better equipped at:

- telling a story with the results of the analysis, and
- crafting a more aesthetic representation of the data plots (or guiding me in doing so).

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.