

# CSC343 Introduction to Databases

## Project Phase One

September 30, 2021

Sarah Mansoor 1004183251 Sumaya Hassen 1000968679

---

# 1 Domain

In this section we will cover the domain we have chosen for our project. We have chosen **education** as our domain.

## 2 Dataset

**A link to our dataset:** [Our Dataset](#)

**Which of the information is relevant to our project:** Board Number, Board Name, Board Language, Board Type, Grade 6 EQAO Reading Results, Progress in Grade 6 EQAO Reading Results, Grade 10 OSSLT Results, Progress in Grade 10 OSSLT Results, Credit Accumulation by the end of Grade 11, Progress in Credit Accumulation by the end of Grade 11, Four Year Graduation Rate, Progress in Four Year Graduation Rate, Five Year Graduation Rate, Progress in Five Year Graduation Rate

**Any learning we will have to do in order to interpret the data:**

- We had to identify what progress meant. The progress number refers to the change in results from 2016-2017 to 2018-2019. We learned this from [here](#)
- Learn more about the domain (Education in Ontario)
- Learn more about the technicalities of SQL to be able to confidently and efficiently write queries to find answers to our questions (AVG, MAX, MIN, etc.).

**Any cleaning we will have to do in order to use the dataset:**

- Check if there any duplicates, confirm whether or not that is okay for our dataset.
- Check if the numerical data adds up and makes sense.
- Check for any spelling errors.
- Check for incorrect data types in any of the columns being used.
- Take out the columns not being used for our project.
- Check for any null or NA values, confirm if they should be replaced or if the row should be deleted.

### 3 Questions

1. Does the average graduation rate differ between the English and french school boards? What about between Roman Catholic and Public school boards?
2. Do the results and progress of EQAO scores vary from the results and progress of OSSLT scores for each board type?
3. Do boards that have a negative progress in credit accumulation by end of grade 11 also have a negative progress in four year graduation rate?

### 4 Schema

#### Relational Schema:

- Board(bID, bName, bLanguage, bType)

A tuple in this relation represents a school board in Ontario. *bID* is the school board's number, *bName* is the school board's name, *bLanguage* is the school board's language, and *bType* is the school board's type. (The possible values for *bLanguage* and *bType* are defined in the integrity constraint below).

- EQAO(bID, eResults, eProgress)

A tuple in this relation represents EQAO information for each school board. *bID* is the school board's number, *eResults* is the EQAO results for the school board, *eProgress* is the progress in EQAO results for the school board. (See the constraints below for more information).

- OSSLT(bID, oResults, oProgress)

A tuple in this relation represents OSSLT information for each school board. *bID* is the school board's number, *oResults* is the OSSLT results for the school board, *oProgress* is the progress in OSSLT results for the school board. (See the constraints below for more information).

- CreditAccumulation(bID, credit, progress)

A tuple in this relation represents credit accumulation information for each school board. *bID* is the school board's number, *credit* is the credit accumulation result by the end of grade 11 for the school board, *progress* is the progress in credit accumulation results by the end of grade 11 for the school board. (See the constraints below for more information).

- Graduation(bID, fourRate, fourProgress, fiveRate, fiveProgress)

A tuple in this relation represents graduation information for each school board. *bID* is the school board's number, *fourRate* is the four year graduation rate for the

school board, *fourProgress* is progress in four year graduation rate for the school board, *fiveRate* is the five year graduation rate for the school board, *fiveProgress* is progress in five year graduation rate for the school board. (See the constraints below for more information).

### Integrity Constraints:

- $\Pi_{bLanguage} \text{ Board} \subseteq \{ \text{"English"}, \text{"French"} \}$
- $\Pi_{bType} \text{ Board} \subseteq \{ \text{"Roman Catholic"}, \text{"Public"} \}$
- $\text{EQAO}[\text{bID}] \subseteq \text{Board}[\text{bID}]$
- $\sigma_{eResults < 0 \wedge eResults > 1} \text{EQAO} = \emptyset$
- $\sigma_{eProgress > 1} \text{EQAO} = \emptyset$
- $\text{OSSLT}[\text{bID}] \subseteq \text{Board}[\text{bID}]$
- $\sigma_{oResults < 0 \wedge oResults > 1} \text{OSSLT} = \emptyset$
- $\sigma_{oProgress > 1} \text{OSSLT} = \emptyset$
- $\sigma_{credit < 0 \wedge credit > 1} \text{CreditAccumulation} = \emptyset$
- $\sigma_{progress > 1} \text{CreditAccumulation} = \emptyset$
- $\sigma_{fourRate < 0 \wedge fourRate > 1} \text{Graduation} = \emptyset$
- $\sigma_{fourProgress > 1} \text{Graduation} = \emptyset$
- $\sigma_{fiveRate < 0 \wedge fiveRate > 1} \text{Graduation} = \emptyset$
- $\sigma_{fiveProgress > 1} \text{Graduation} = \emptyset$

### Data dictionary:

**Board**(bID, bName, bLanguage, bType)

Attribute	Description	Type	Known Value	Default Value	Allowable Values
<u>bID</u>	Board Number	INT	Yes	-	B#####
bName	Board Name	TEXT	Yes	-	-
bLanguage	Board Language	TEXT	Yes	English	English, French
bType	Board Type	TEXT	Yes	Public	Roman Catholic, Public

Note: B##### represents the letter B followed by 5 numbers. Ex: B11111.

### **EQAO(bID, eResults, eProgress)**

Attribute	Description	Type	Known Value	Default Value	Allowable Values
<u>bID</u>	Board number	INT	Yes	-	B#####
eResults	EQAO Results	NUM	Yes	-	$0 < \text{eResults} < 1$
eProgress	EQAO Progress	NUM	Yes	-	$\text{eProgress} < 1$

### **OSSLT(bID, oResults, oProgress)**

Attribute	Description	Type	Known Value	Default Value	Allowable Values
<u>bID</u>	Board number	INT	Yes	-	B#####
oResults	OSSLT Results	NUM	Yes	-	$0 < \text{oResults} < 1$
oProgress	OSSLT Progress	NUM	Yes	-	$\text{oProgress} < 1$

### **CreditAccumulation(bID, credit, progress)**

Attribute	Description	Type	Known Value	Default Value	Allowable Values
<u>bID</u>	Board number	INT	Yes	-	B#####
credit	Credits done	NUM	Yes	-	$0 < \text{credit} < 1$
progress	Progress in credits done	NUM	Yes	-	$\text{progress} < 1$

### **Graduation(bID, fourRate, fourProgress, fiveRate, fiveProgress)**

Attribute	Description	Type	Known Value	Default Value	Allowable Values
<u>bID</u>	Board number	INT	Yes	-	B#####
fourRate	4-yr Graduation Rate	NUM	Yes	-	$0 < \text{fourRate} < 1$
fourProgress	4-yr Graduation progress	NUM	Yes	-	$\text{fourProgress} < 1$
fiveRate	5-yr Graduation Rate	NUM	Yes	-	$0 < \text{fiveRate} < 1$
fiveProgress	5-yr Graduation progress	NUM	Yes	-	$\text{fiveProgress} < 1$

### **Justification of design:**

Similar columns that had related information were grouped together in a table. Thus, the existing relations consist of board information, EQAO scores and progress, OSSLT scores and progress, Credit Accumulation, and Graduation rates and progress. Grouping related information will make it easier to query. We also decided for the *bID* to be a

foreign key for the other relations. That way we could connect relations such as EQAO, OSSLT, Credit Accumulation, and Graduation to their respective school boards. This also helps us compare the relations.

Now we'll explain why we designed each relation the way we did:

Board(bID, bName, bLanguage, bType): We created this relation to identify information about the boards we explore. Including the *bID* gives us a way to uniquely identify each board as well as *bName*, the board name. The *bLanguage* allows us to write query's for when we explore the differences or similarities between English and French school boards as does *bType* to explore differences between Roman Catholic and Public school boards.

EQAO(bID, eResults, eProgress) and OSSLT(bID, oResults, oProgress) and CreditAccumulation(bID, gID, credit, progress): EQAO, OSSLT and CreditAccumulation are designed similarly. As we stated before, *bID* allows us to connect relations from EQAO, OSSLT and CreditAccumulation results and progress to their respective school boards. CreditAccumulation also includes *gID*, the grade, for which the credit results are from (ie. grade 10 or grade 11).

Lastly, Graduation(bID, fourRate, fourProgress, fiveRate, fiveProgress): *bID* allows us to connect relations from graduation results and progress to it's respective school boards. We want to compare four year graduation and five year graduation which is why we included *fourRate*, *fourProgress*, *fiveRate*, *fiveProgress*. Additionally, *fourProgress* and *fiveProgress* will allow us to see the differences between the progress of four and five year graduation rates.