

# Project Documentation

Sean Hambali

December 13, 2023

[\[Link to dashboard\]](#)

## Executive summary

The COVID-19 pandemic has caused significant student learning losses throughout D.C. Using a descriptive tabulation on statewide assessment data from D.C.'s Office of the State Superintendent of Education (OSSE), I show that in the immediate aftermath of the COVID-19 pandemic, proficiency in the English and Language Arts (ELA) subject declines by 6.3 percentage points, eradicating nearly 1.8 years' worth of learning gains progress. The decline is even steeper for Math subject, where proficiency declines by 10.4 percentage points, or equivalent to 5.4 years of learning gains. Tabulations across subgroups show that lower-income areas, vulnerable population groups, and early grade students face disproportionate learning losses. To aid in mitigating learning losses, this dashboard provides a school-level performance tracker that allows the identification of schools that are in most need for policy assistance.

## Context

The first COVID-19 case in D.C. was recorded on March 7, 2020. As with other administrations all around the world, the government had enacted various restriction measures that were aimed at curbing the spread of the virus. Among the most consequential measures taken were the decision to shutdown schools and to transition teaching-learning activities to a fully-remote environment.

This measure resulted in tremendous learning losses (Goldhaber et al., n.d.). A study estimates that the average U.S. public school student in grades 3-8 has lost an equivalent of half year of learning in math, and a quarter of a year in reading (Fahle et al. 2023). Moreover, it is likely that the pandemic is affecting students differently. Students from well-off households typically have more resources to support at-home learning; these students are more likely to be able to afford proper digital infrastructure, more learning resources at home, and they are more likely to study at schools that are better positioned to navigate the transition to at-home learning.

In this dashboard, I intend to provide visual and descriptive explorations of whether, and to what extent student learning losses vary across student groups and areas in D.C. The

results can be used to aid policymakers in focusing policy interventions on groups and/or areas that are hardest hit by remote instruction policies, and are most lagging in terms of learning recovery.

## **Visualization goals and intended audience**

When developing this dashboard, I have two users in mind: (1) layman who are interested in the education policy sector; (2) policymakers in the education sector who need information on the list of schools to be targeted for policy interventions.

The goals of this visualization are as follows:

1. To communicate the magnitude of learning losses in D.C. following the COVID-19 pandemic
2. To visualize the importance of socioeconomic factor in explaining the magnitude of learning loss throughout D.C.
3. To identify schools that are in greatest need for relevant policy interventions

To ensure that the analysis is accessible to layman that have rudimentary background in statistics, the visualizations that are presented in the dashboard only include simple tabulations of percentages, albeit across different groups. A metadata is also present below every chart to make more explicit the data sources for each chart, and the processing steps that are done to create the tabulations.

The page in the dashboard that might be of particular interest to policymakers is the third tab, i.e. “Tracking learning loss at the school-level”. The tab contains information on the list of schools that users can play around with interactively, sorting schools based on wards, magnitudes of proficiency declines, among others.

## **Methods and data description**

All data in the dashboard are sourced from D.C.’s Office of the State Superintendent of Education (OSSE), taken from the publicly accessible yearly school-level performance report card. Every year, D.C. implements statewide assessment to assess students’ college readiness via Partnership for Assessment of Readiness for College and Careers (PARCC), or the Multi-State Alternative Assessments (MSAA) for students with learning disability. These PARCC and MSAA scores are then reported to the OSSE for public reporting purposes. Hence, to the best of my knowledge, there isn’t any procedures that need to be followed in regards to data privacy or data protection. I use data from years 2016, 2017, 2018, 2019, 2022, and 2023. Data from years 2020 and 2021 are missing, since PARCC/MSAA assessments were not implemented during at-home learning period.

The main indicator that I use for the tabulations in the dashboard is the PARCC/MSAA proficiency rate, which is simply just the percentage of test takers that meet or exceed expectations, defined as having scores 4 or more in the assessment’s 1-5 scale. The alternative measure that I use is the difference in proficiency rates between 2022 and the 2016-2019 average, which depicts the magnitude of departure from baseline average.

Table 1: Snippet of first 10 rows in the OSSE 2022 proficiency data

School Code	Assessment Name	Subject	Student group	Subgroup Value	Total Count	Percent
1058	PARCC	Math	Race/Ethnicity	Hispanic/Latino	n<10	n<10
1058	All	Math	Race/Ethnicity	Black/African American	DS	DS
1058	All	Math	Special Programs	Active or Monitored English Learner	n<10	n<10
1058	PARCC	Math	All	All	DS	DS
1058	All	Math	Race/Ethnicity	Black/African American	n<10	n<10
1058	PARCC	Math	Special Programs	Not Active or Monitored English Learner	DS	DS
1058	All	ELA	Special Programs	Students with Disabilities	n<10	n<10
1058	PARCC	ELA	Special Programs	Not Military Connected	202	27.72
1058	PARCC	ELA	Special Programs	English Learner	n<10	n<10
1058	All	Math	Special Programs	Not CFSA	n<10	n<10

There are several challenges in working with the OSSE data. First, the data is not reported in a standardized format across the years. It is noteworthy that each row in the OSSE data does not reference schools or students; rather, each row represents a summary of a single dimension, as shown by Table 1.

One problem with this is that the subgroup value is not harmonized across the years. Some years might have a differentiation by gender, while others might not. To address this issue, I first harmonize the levels of the subgroups of interest. Regardless, this presents the analysis with an important limitation; while some dimensions, such as student gender, might be important in explaining the heterogeneity in learning loss, they are not included in the dashboard because there isn't enough year to display meaningful variation along these dimensions. Further, the numeric columns in some years contain “%” while absent in others, so I also harmonize these by using regular expressions to only retrieve the numeric components of the numeric variable.

Another challenge that can be observed from Table 1 is the fact that some dimensions do not have enough student counts (below the minimum reporting threshold of 10), and as such, the proficiency rates are sometimes suppressed. Another non-valid entry is the DS code, which means that the number was secondarily suppressed to ensure that other suppressed data cannot be determined. I regard all these non-valid entries as missing and therefore omit them from my analysis.

The result of the cleaning is a school-subject dataframe that contains the count of valid test takers and the proficiency rates, both by aggregate and also by student subgroups. Whenever an aggregation is required, I compute a weighted average of the proficiency rate, using the count of valid test takers as the weights.

I create the dashboard using a combination of Shiny App in R and Plotly tools that allow users to interactively play around with the charts or tables in the dashboard.

## Analysis and key insights

I use descriptive tabulations to achieve the above-mentioned visualization goals. First, I summarize the by-subject yearly variations in overall proficiency scores (aggregated across all D.C.). Second, I explore the heterogeneities in learning loss along several socioeconomic dimensions, such as median household income, among others. Lastly, I also calculate the

score difference at the school level as a rough measure of the school’s learning loss magnitude (or lack thereof) .

The key insights from the dashboard were as follows:

1. **There have been widespread losses in student learning** across D.C. The descriptive tabulation shows that proficiency in ELA decreases by 6.3 percentage points from 2019 to 2022, eradicating **1.8 years of learning gain progress**. Further, losses are even steeper for Math subject, where proficiency declines by 10.4 points, or equivalent to **5.4 years of learning gains**.
2. However, **not every students is equally affected**. There is a strong socioeconomic aspect to the learning loss issue. Students that are **of lower-income families, are economically-disadvantaged, and are Black** faced disproportionate learning losses.
3. Worryingly, learning losses, especially in Math, tend to be concentrated among **early primary students** (grades 3 and 4), which could potentially jeopardize long-term human capital.
4. This dashboard aids policymakers in **identifying schools with the greatest need for policy intervention**.
5. One promising avenue of intervention is **tutoring** (Nickow, Oreopoulos, and Quan 2020). When properly implemented, tutoring programs have been shown to consistently improve student learning outcomes, with an effect size of 0.37 standard deviations – this roughly translates to a student’s advancement from the 50th to 66th score percentile.
6. Policymakers should, then, consider designing a large-scaled tutoring program led by teachers or paraprofessionals. In addition, there is large value in focusing effort on students of earlier grades, where, in our case, steepest learning losses tend to occur, and also where the program can potentially yield the largest impacts.

## Bibliography

- Fahle, Erin, Thomas Kane, Tyler Patterson, Sean Reardon, Douglas Staiger, and Elizabeth Stuart. 2023. “New Data Show How the Pandemic Affected Learning Across Whole Communities | Harvard Graduate School of Education.” <https://www.gse.harvard.edu/ideas/news/23/05/new-data-show-how-pandemic-affected-learning-across-whole-communities>.
- Goldhaber, Dan, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger. n.d. “The Consequences of Remote and Hybrid Instruction During the Pandemic.” <https://doi.org/10.3386/w30010>.
- Nickow, Andrew Joshua, Philip Oreopoulos, and Vincent Quan. 2020. “The Transformative Potential of Tutoring for Pre K-12 Learning Outcomes: Lessons from Randomized Evaluations.” <https://www.povertyactionlab.org/publication/transformative-potential-tutoring-pre-k-12-learning-outcomes-lessons-randomized>.