The Impact of Pandemic Learning Models In Pennsylvania: Assessing Educational Outcomes Using Matching

Mena Tetali, Sara Murley, and Samantha Rudravajhala

November 2024

Abstract

The COVID-19 pandemic greatly disrupted education systems, raising questions about the impact of varying school reopening policies. This study examines whether Pennsylvania school districts with similar characteristics but different COVID-19 policies, specifically whether they offered in-person instruction during the 2020-2021 school year, experienced different educational outcomes in the subsequent year. Using district-level data, we matched districts with comparable characteristics that differed in their degree of in-person instruction and analyzed standardized test scores before and after the pandemic. Our findings reveal that districts with more in-person instruction during the studied period experienced significantly smaller declines in Math test scores, with an average difference of 0.97 percentage points ($t = 2.54, df = 188$) compared to districts without in-person instruction. These results highlight the potential academic benefits of in-person learning amidst unprecedented disruptions.

Keywords— Pandemic Learning Models, Pennsylvania, Educational Outcomes, Matching

1 Introduction

In 2020, the COVID-19 pandemic swept the United States, forcing schools nationwide to quickly adapt to new methods of instruction. The most common approach was remote or online learning, a shift that posed significant challenges for both teachers and students. Many students struggled to maintain focus and were less likely to seek help when they encountered difficulties, exacerbating learning gaps. In the years following the pandemic, studies have documented declines in student test scores, underscoring the lasting impact of this disruption.

The effects of the pandemic spared no one: individuals of all ages and demographics were affected in some way. However, students proved to be particularly vulnerable, as recent studies indicate a significant decline in learning quality during this period. The abrupt shift to less immersive and engaging methods of instruction in March 2020 disrupted the traditional classroom experience, hindering students' ability to interact effectively and fully engage with their education. Research on the educational impact of COVID-19 emphasizes the need for targeted, student-centered learning that leverages the benefits of both synchronous and asynchronous instruction [12]. Additionally, the United Nations' 2020 policy brief underscored how the pandemic exacerbated existing disparities in access to education, particularly for underserved communities such as refugees, low-income families, and individuals with disabilities [8]. However, the brief also highlighted how the crisis has driven innovation in the education sector, underscoring the importance of delivering high-quality, engaging education through diverse modalities to meet the needs of all demographics.

In Pennsylvania, Governor Tom Wolf issued a statewide shutdown of all 'non-life-sustaining businesses and services', including schools, on March 19, 2020. Stay-at-home orders would soon be implemented, and students adapted to a "new normal" through virtual learning. Throughout the pandemic, however, differences in COVID-19 in person instruction policies between school districts significantly shaped student performance across the state. These discrepancies in learning models would become evident in the academic outcomes on the Pennsylvania System of School Assessment exam observed in the following school year.

With our research, we aim to answer the question: **How did Pennsylvania districts with any in-person instruction during the 2020-2021 COVID-19 school year compare to fully remote districts in post-pandemic state test scores?**

To understand this relationship, we will analyze district level data in the Commonwealth of Pennsylvania, looking at how long districts were in-person during the 2020-2021 school year. We employ matching to create pairs of districts that share certain characteristics but differ in their COVID-19 policies in order to investigate differences in test scores before and after the pandemic between district pairs.

2 Data and Methods

This project uses data from multiple online sources: Future Ready PA, Digital Bridge K-12, COVID-19 School Data Hub, and the Commonwealth of Pennsylvania database:

2.1 Future Ready PA

Future Ready PA is an index of measures related to student and school success. It contains a collection of school progress measures to "more accurately report student learning, growth, and success in the classroom and beyond" [10]. The website contains datasets listed as District Fast Facts and School Fast Facts. For the purpose of our research, we used the District Fast Facts from the 2023-2024 school year. This dataset contains variables such as percentage of economically disadvantaged students, percentage of gifted students, etc. This dataset was crucial to our research because it allowed us to match districts with similar demographics.

2.2 Digital Bridge K-12

Digital Bridge K-12 is an interactive map that documents the number of unconnected students across the country[3]. Unconnected students means students who don't have internet access at home. We were able to access a dataset that has an estimate of the number of unconnected students in each school district within Pennsylvania. This dataset allowed us to add internet connectivity to the list of covariates that we had obtained from Future Ready PA.

2.3 COVID-19 School Data Hub

The COVID-19 School Data Hub is "a central database for educators, researchers, and policymakers to understand how the COVID-19 pandemic shaped students' modes of learning in 2020-21."[7] The database tracked whether a Pennsylvania school district was virtual, hybrid, or in-person for each month of the 2020-2021 school year. We aggregated the number of months spent in-person across each district to assess the causal effect of a district's COVID-19 policy on educational outcomes.

2.4 Commonwealth of Pennsylvania

We obtained standardized test scores through the Commonwealth of Pennsylvania website[11]. To properly measure the impact that the pandemic had on educational outcomes, we decided to use test score data from the Pennsylvania System of School Assessment (PSSA) exam both before and after the pandemic. The PSSAs are an annual assessment administered in all Pennsylvania schools from grades 3-8. We utilized test score data for 2019 (pre-pandemic) and 2022 (post-pandemic). In order to clean this data to be used in our analysis, we calculated the difference in the average percentage of students in each district who scored proficient in English and Math before and after the pandemic.

2.5 Data Wrangling and Final Dataset

After we had gathered our data from these various sources, we performed various data cleaning techniques and merged them together into one dataset on which we could perform our matching analysis. The resulting data set included one row per district, with covariates describing demographic characteristics, aggregate time spent online, in person, and online, and percentage of proficient scores in both English and Math. Table 1 includes descriptive statistics for each of the test-score variables in our final dataset.

Variable	**Mean**	**St. Dev**	**Min**	**Max**	**Obs**
2019 % English Proficiency	45.617	5.989	14.2	57.1	432
2019 % Math Proficiency	28.651	6.195	2.75	40.02	432
2022 % English Proficiency	41	6.525	12.1	53.1	432
2022 % Math Proficiency	24.439	6.765	2.5	37.02	432

Table 1: Sample descriptive statistics

3 Analysis

In order to analyze the impact of COVID-19 learning models on educational outcomes, we explored various tools and methods. We began by defining a treatment variable to be our policy of interest. We used a histogram 1 to examine the distribution of months spent in person across the school districts in our dataset. The distribution is strongly skewed to the right, with very few districts spending any time in person. The median of the distribution is 1 month spent in person. We decided to define our treatment variable using the median value to separate the groups, as below:

$$\text{Treatment} = \begin{cases} 1 & \text{Districts that spent at least 1 month fully in person during the examined period} \\ 0 & \text{Districts that spent 0 months in person during the examined period} \end{cases}$$

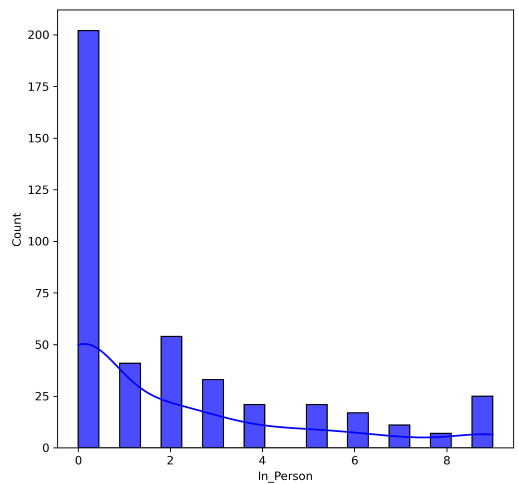

Figure 1: Histogram of Months in Person

After defining the treatment variable, we also decided which specific outcomes we wanted to analyze to assess educational outcomes. The test scores dataset that we used included various variables by district, though we elected to look at the mean number of students that scored proficient on the English and Math exams in the schools within each district. We chose to exclude the science section as it is not administered across all grades 3-8. We therefore defined two outcome variables that we were interested in examining at the district level: the difference in the average number of students scoring proficient in English from 2019 to 2022, and the difference in the average number of students scoring proficient in Math from 2019 to 2022.

In our analysis, we used the following covariates: District Enrollment, % Gifted Students, Charter School Enrollment, Geographic Size, Demographic Percentages (American Indian, Asian, Native Hawaiian or Pacific Islander, Black, Hispanic, White, Two or More Races, Female, Male), Economically Disadvantaged, English Learners, Special Education, Homeless, Military Connected, and Percentage of Unconnected Students. Each of these district characteristics presumably plays a large role in the test scores of the district and potentially in the COVID-19 learning models that the district selected. By controlling for these covariates, we aim to establish causality between the learning models and the educational outcomes. Furthermore, before performing any statistical analysis or training any machine learning models, it is important to normalize each of the numerical covariates. To do so, we used scikit-learn's standard scaler.

Since our data is observational and randomization through an experiment is not possible in the context of this analysis, we looked for methods that would allow us to establish a causal effect of the treatment variable on the outcomes. We began by running simple OLS regression models on both of our outcome variables separately (difference in English test scores and Math test scores) as a function of the treatment variable and all of the covariates. The results from the regression are shown in Table 3 in the results section. However, we were also interested in exploring other techniques to assess the robustness of our results.

As an alternative to establishing causality through controlling for covariates in an OLS model, we explored matching. Matching is a statistical technique that compares treatment and control groups to evaluate the effect of the treatment on an outcome variable. The method aims to reduce biases by creating treatment and control comparison groups that are similar according to a set of covariates. This process helps to ensure that any differences in outcomes between the treatment and control groups are attributable to the treatment itself, not to differences in other variables [9, Chapter 31].

Matching presents certain advantages and disadvantages as compared to other statistical methods that we considered for our analysis. Ho et al., for example, argue that matching avoids the dependence of accurate results on the exact specification of the model (defining covariates in a linear regression model) [6]. Matching is especially useful in circumstances where the exact reasoning behind the assignment of the treatment and control groups is unknown, as is true in our case [1]. Furthermore, matching is less restrictive than an OLS model, as there are no assumptions about the functional form of the relationship between the independent and dependent variables (i.e. linearity) [2]. The exact matching approach that we used is matched difference-in-differences [1].

The process involves two key steps:

1. Match treatment to control districts that are similar across covariates
2. Use a t-test to compute the difference of means across pairs to assess significance

The first tool we used for matching was the nearest neighbors function as part of the scikit-learn library. We separated the dataset into treatment and control groups, then used the nearest neighbors function that minimized the Euclidean distance metric to create pairs across the two groups. However, this method did not result in a one-to-one matching, instead having some school districts appearing multiple times in the matched dataset but paired with various districts. As such, we decided to explore other tools that are designed more specifically for the matching process.

After exploring multiple libraries[1], we elected to use the MatchIt library in R. The library encompasses several matching methods. The most commonly used is the nearest neighbor method with propensity score difference as the distance metric [5]. Propensity score is calculated using logistic regression. The method works by finding the closest eligible control unit for each treatment unit according to the defined distance metric. This differs from other matching techniques that aim to optimize an overall criterion across all pairs. With propensity score as the distance metric, the default order for matching is to begin with the points that would have the hardest time finding a close match. This method also allows for matching without replacement, including a strategy for ties to be broken so that the resulting dataset contains a one-to-one matching [5].

A summary of the matching specification and the process used for arriving at it can be found in the below table (obtained using the print method on the MatchIt object in R):

Method	1:1 nearest neighbor matching
Replacement	FALSE
Distance Metric	Propensity Score
Number of Observations (Original)	432
Number of Matched Pairs	378

The results from the implementation of the matching in R contains a dataset with one row per school district with the following additional variables (not included in the inputted data set of covariates, treatment and outcomes):

- ID: ID of the school district used for pairing
- Subclass: pair ID (matched school districts will share the same value)
- Distance: individual propensity score

The MatchIt library also includes methods to assess and report the quality of a matching specification [4]. When balance is achieved, the resulting effect is less likely to be influenced by model misspecification and ideally close to the true

[1]The PyMatch library is not compatible with updated versions of Pandas, though presumably would have worked well for our analysis.

effect [4]. The summary method provides various balance metrics for the matched data across covariates, which we consolidated into Table 2. We will first report standardized mean differences of each covariate before and after the matching. SMDs close to 0 indicate good balance. The difference in SMDs before and after matching is shown in Figure 2 below, revealing that the matching process significantly improved balance. We also report the variance ratios between the treatment and control groups for each covariate. Variance ratios close to 1 indicate balanced groups, though a commonly recommended range is between 0.5 and 2 [4]. Finally, we include the eCDF statistics, or the overall difference in distributions of each covariate between the groups. Values closer to 0 indicate better balance. Based on the results below, we can conclude that our matching specification resulted in balanced treatment and control groups and we can proceed with the statistical analysis.

Variable	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max
Number of Schools	4.4550	4.5608	-0.0449	0.6573	0.0138	0.0423
District Enrollment	-0.0967	-0.0801	-0.0501	0.7186	0.0229	0.0794
Percent of Gifted Students	0.1411	0.0168	0.1127	1.3401	0.0244	0.0741
Charter School Enrollment	-0.0664	-0.0637	-0.1093	0.6906	0.0187	0.0635
Geographic Size of District	0.1501	0.0603	0.0903	0.7702	0.0732	0.1852
American Indian/Alaskan Native	0.1136	-0.0714	0.1214	13.2690	0.0243	0.0582
Asian	-0.0764	-0.0396	-0.0449	0.8112	0.0173	0.0529
Native Hawaiian or Other Pacific Isl.	-0.0552	0.0298	-0.0949	0.8595	0.0112	0.0423
Black	-0.2641	-0.2239	-0.0902	1.2324	0.0291	0.0688
Hispanic	-0.1697	-0.1236	-0.0799	0.7064	0.0167	0.0635
White	0.2967	0.2274	0.1129	0.8618	0.0274	0.0688
Two or More Races	-0.1950	-0.1031	-0.1213	0.8218	0.0244	0.0794
Economically Disadvantaged	-0.1905	-0.1305	-0.0724	0.8749	0.0388	0.1534
English Learner	-0.2084	-0.1643	-0.0775	0.8813	0.0190	0.0582
Special Education	-0.1560	0.0219	-0.1928	0.8834	0.0441	0.1270
Female	0.0187	-0.0218	0.0374	1.2411	0.0174	0.0741
Male	-0.0218	0.0220	-0.0404	1.2391	0.0177	0.0688
Homeless	-0.1356	-0.0512	-0.1110	0.5588	0.0181	0.0688
Military Connected	0.0906	0.0067	0.0852	0.8291	0.0277	0.0952
perc_unconnected_students	-0.2500	-0.1245	-0.1654	0.7550	0.0335	0.1164
Latitude	40.6583	40.6581	0.0004	1.1727	0.0372	0.1005
Longitude	-77.6069	-77.7211	0.0644	0.9200	0.0262	0.0794

Table 2: Comparison of Means, Standardized Differences, and Other Metrics

The resulting data frame of matches was exported to a .csv file to be used in the remainder of the analysis in Python. The data frame was consolidated and pivoted to include a row for each pair, with a column for the index of the treatment district and a column for the index of the control district within the pair. A unique challenge to this analysis came from the different indexing strategies used by R and Python. To combat the fact that R begins indexing on 1, an empty row was added to the beginning of the dataframe in Python.

To perform a difference of means t-test, we first calculated the difference between the outcome variables for each district in the pair. We then used the ttest-rel method within scikit-learn to perform a paired t-test. The results of the test are

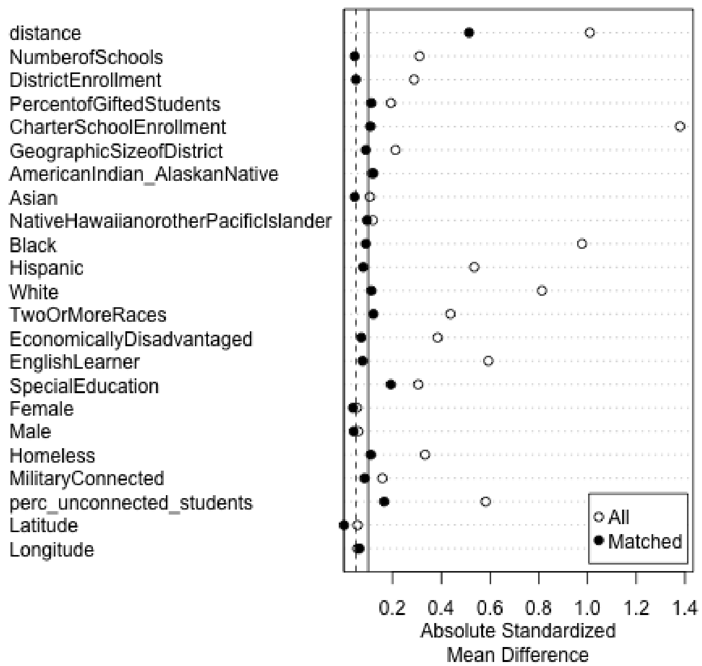

Figure 2: Standardized Mean Differences Before and After Matching

shown and discussed in the results section below.

4 Results

The results from the regression analysis are shown in the Table 3 below. For simplicity, only the treatment and constant are included, though a complete regression table with all covariates can be found in Appendix A. The table includes results from two models with identical independent variables, but the following dependent variables:

- Change in average number of students scoring proficient in English from 2019 to 2022
- Change in average number of students scoring proficient in Math from 2019 to 2022

The regression results for the independent variable of interest, treatment, differ across models. For English, schools with the treatment policy had an average decrease in mean proficiency of 0.67 percentage points less than the control group. The results are not statistically significant at the 0.05 significance level (p = 0.09, df = 408). For Math, schools with the treatment policy had an average decrease in mean proficiency of 1.03 percentage points less than the control group. The result is significant at the 0.01 significance level (p = 0.01, df = 408).

	Difference in English Proficiency	Difference in Math Proficiency
Treatment	0.67	1.03
(SE)	(0.39)	(0.38)
[p]	[0.09]	[0.01]
Constant	-11.86	-5.91
(SE)	(17.78)	(17.08)
[p]	[0.51]	[0.73]
N	423	423

Table 3: Regression Results

We compare these results to the results from our difference of means t-test on our matched pairs. The t-test results are shown in Table 4 below. 95% confidence intervals for the mean differences are shown in Figure 3. On average, schools with the treatment policy had 0.66 percentage points less of a decrease in English proficiency as compared to their paired schools in the control group. The results are not statistically significant at the 0.05 significance level ($t = 1.81, p = 0.071, df = 188$). On average, schools with the treatment policy had 0.97 percentage points less of a decrease in Math proficiency as compared to their paired schools in the control group. The results are statistically significant at the 0.05 significance level ($t = 2.54, p = 0.012, df = 188$).

Test	**Statistic**	**P-Value**	**Degrees of Freedom**
English T-Test	1.81	0.071	188
Math T-Test	2.54	0.012	188

Table 4: T-Test Results for English and Math Test Scores

To ensure robustness, or the consistency and reliability of our results across different methods and models, we employed both OLS regression and a difference-

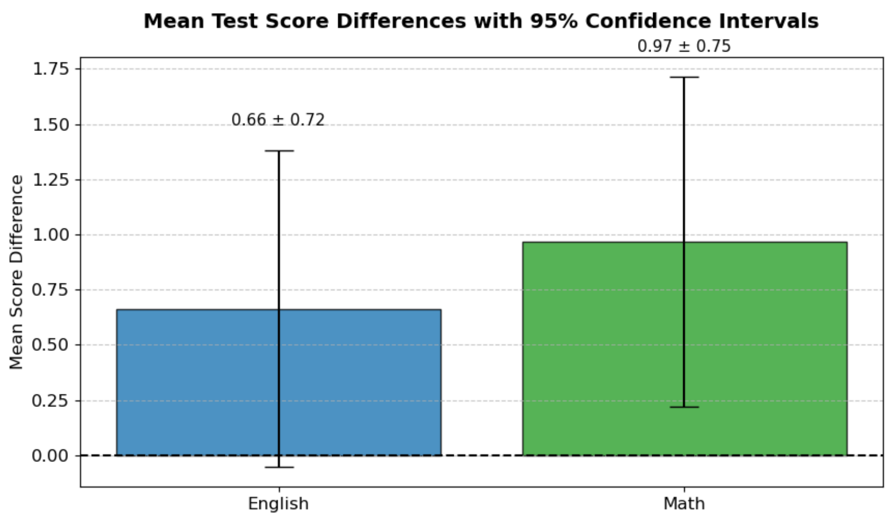

Figure 3: Confidence Intervals for Difference of Means

of-means t-test on matched pairs. By using multiple approaches, we are able to confirm that our findings are robust and more likely representative of a true relationship. For both English and Math, the treatment effect magnitudes are similar across the OLS regression and the matched t-test result (0.67 vs. 0.66 for English and 1.03 vs 0.97 for Math). This similarity suggests that both methods capture a consistent treatment effect across models. Furthermore, the OLS model for English proficiency shows slight significance with a p-value of 0.09, while the matched t-test result is slightly stronger with a p-value of 0.071. However, neither are statistically significant at the 5% level. Both methods for Math proficiency, on the other hand, yield significant results at the 5% level, with the OLS model ($p = 0.01$) showing slightly stronger evidence than the matched t-test ($p = 0.012$).

Methodological differences between these two methods are important to highlight when assessing these results. The OLS regression controls for covariates directly and assumes a linear relationship between predictors and outcomes. Another limitation is that it relies on the assumption that the model is correctly specified in the linear form and that covariates adequately control for endogeneity. In contrast, the matched pairs t-test balances treatment and control groups by pairing schools with similar demographics (covariates). This approach reduces the reliance on model assumptions, however, it does not explicitly control for unmeasured variables that might still introduce bias.

Both methods present different strengths and weaknesses, but the consistency of the treatment effects across both models in terms of magnitude and statistical significance suggests that these findings are robust. The effects of the treatment policy on English and Math proficiency are consistent in direction and size, while the statistical significance for Math proficiency across both methods confirms a negative effect. For English proficiency, the lack of statistical significance in both approaches suggests a weaker and less definite effect.

Overall, the results for the treatment effect on the difference in the percentage of students proficient in Math is robust across methods, showing a significant effect of the treatment policy. For English proficiency, the treatment effect is consistent in magnitude but remains statistically insignificant for both models.

5 Discussion

Ultimately, our results indicate that in Pennsylvania, the mode of learning played a critical role in determining student performance after the COVID-19 pandemic. Our findings reveal that in-person learning was a significant factor in explaining variations in academic outcomes among students in similar districts. Using matching and OLS techniques, we found that districts that were in-person for longer during the 2020-2021 school year experienced a smaller decline in scores, especially in Math, highlighting the importance of immersive learning environments. These results suggest that face-to-face instruction can mitigate the negative effects of disruptions on student performance. Moving forward, the evidence underscores the potentially imperative role of in-person and immersive learning in safeguarding academic achievement during future crises. Policymakers and educators should prioritize strategies to maintain or safely facilitate in-person learning where possible, as it can be instrumental in reducing the adverse impact on student outcomes in key subject areas.

The next steps for this analysis involve conducting a deeper investigation into how learning models have impacted different demographics, particularly when it comes to COVID-19 policies and educational outcomes. A key question to ask is whether these policies have had varying effects on different pairs. For example, do the policies have more of an impact for pairs of low-income districts or those with lower internet access? Understanding these dynamics can provide more nuanced insights into the equity of policies during disruptions and would help identify populations in need of more targeted interventions.

Additionally, experimenting with different analytical approaches such as different matching methods or spatial mapping techniques, will be important to assessing the robustness of the results. These new methods can help refine comparisons. Further, determining the impact of various learning models on test score differences across different states will help us understand the broader implications of educational disruptions in the United States. By taking these steps, we can better identify areas for improvement and inform strategies to support equitable learning opportunities in future crises.

References

[1] The World Bank, *Matching - dimewiki*, https://dimewiki.worldbank.org/Matching, 2024.

[2] Dan A. Black, *Matching as a regression estimator*, IZA World of Labor **186** (2015), 1–10.

[3] EducationSuperHighway, *Interactive map: America's unconnected students*, https://digitalbridgek12.org/toolkit/assess-need/connectivity-map/, 2024.

[4] Noah Greifer, *Assessing balance*, https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html, 2024.

[5] ———, *Matching methods*, https://cran.r-project.org/web/packages/MatchIt/vignettes/matching-methods.html#nearest-neighbor-matching-method-nearest, 2024.

[6] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart, *Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference.*, Political Analysis **15** (2007), no. 3, 199–236.

[7] COVID-19 School Data Hub, *District-monthly percentage in-person, hybrid, or virtual*, https://www.covidschooldatahub.com/data-resources, 2024.

[8] United Nations, *Policy brief: Education during covid-19 and beyond*, https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf, 2020.

[9] Mike Nguyen, *A guide on data analysis*, Bookdown, 2020.

[10] Pennsylvania Department of Education, *Future ready pa index*, https://futurereadypa.org/district/fastfacts?id=069165241186158253181125090032179143149036203064#, 2024.

[11] Commonwealth of Pennsylvania, *Assessment reporting*, https://www.pa.gov/agencies/education/data-and-reporting/assessment-reporting.html, 2024.

[12] Y. Zhao and J. Watterston, *The changes we need: Education post covid-19*, Journal of Educational Change **22** (2021), 3–12.

A Regression Table

Table 5 contains regressions results including all covariates for both models (English and Math test scores).

Table 5: Regression Results for Math and English Difference

Variable	**English Difference**				**Math Difference**			
	Coef.	**Std. Err.**	**t-value**	**p-value**	**Coef.**	**Std. Err.**	**t-value**	**p-value**
(Intercept)	−11.860	17.783	−0.667	0.5052	−5.908	17.075	−0.346	0.7295
treatment	0.669	0.392	1.704	0.0891	1.027	0.377	2.726	0.0067
Number of Schools	0.006	0.133	0.045	0.9643	0.005	0.127	0.040	0.9679
District Enrollment	−0.488	1.141	−0.427	0.6695	0.275	1.096	0.251	0.8023
Percent of Gifted Students	0.527	0.248	2.130	0.0338	0.431	0.238	1.812	0.0707
Charter School Enrollment	−0.921	2.891	−0.318	0.7503	−1.616	2.776	−0.582	0.5609
Geographic Size of District	0.252	0.216	1.165	0.2447	0.401	0.208	1.932	0.0541
American Indian/Alaskan Native	−0.544	0.741	−0.733	0.4637	−0.679	0.712	−0.954	0.3408
Asian	−3.635	10.561	−0.344	0.7309	−7.626	10.140	−0.752	0.4525
Native Hawaiian/Pacific Islander	0.030	0.322	0.093	0.9260	−0.232	0.309	−0.751	0.4530
Black	−12.082	31.745	−0.381	0.7037	−23.763	30.482	−0.780	0.4361
Hispanic	−11.722	28.045	−0.418	0.6762	−21.632	26.929	−0.803	0.4223
White	−20.808	52.645	−0.395	0.6929	−40.326	50.550	−0.798	0.4255
Two or More Races	−3.694	8.919	−0.414	0.6790	−6.859	8.564	−0.801	0.4237
Economically Disadvantaged	0.085	0.475	0.180	0.8573	−0.553	0.456	−1.213	0.2257
English Learner	0.653	0.418	1.561	0.1194	0.087	0.402	0.215	0.8295
Special Education	0.076	0.220	0.347	0.7289	−0.038	0.211	−0.182	0.8557
Female	−5.632	8.584	−0.656	0.5122	−6.894	8.243	−0.836	0.4034
Male	−5.438	8.579	−0.634	0.5265	−6.966	8.237	−0.846	0.3982
Homeless	0.030	0.223	0.132	0.8949	0.196	0.215	0.915	0.3609
Military Connected	−0.223	0.187	−1.191	0.2343	0.116	0.180	0.645	0.5192
Percent Unconnected Students	−0.913	0.470	−1.945	0.0524	0.149	0.451	0.331	0.7406
Latitude	−0.381	0.363	−1.049	0.2950	−0.146	0.349	−0.420	0.6750
Longitude	−0.288	0.132	−2.186	0.0294	−0.092	0.126	−0.724	0.4698
Residual Std. Error	3.728				3.580			
R-Squared	0.1606				0.1111			
F-Statistic (df=23,408)	3.395 (p=3.8e-07)				2.217 (p=0.0011)			