

Schooling Methods in Student Test Scores*

Denise Chang

April 15, 2024

An increasing number of schools and universities have adopted virtual or hybrid teaching methods ever since the COVID-19 pandemic in 2020. Using data from the National Center for Education Statistics and from various district-level assessments, this paper investigates the impact of schooling modes on students' pass rate in state standardized assessments in grades 3-8 during the 2020-2021 school year. The exploration of the data across 11 states suggests that the overall student pass rates declined during the pandemic school year. The pass rates have also seen more drastic changes in schools with larger share of black students as well as in schools who had a higher share of virtual and hybrid schooling. The results of this study are significant as they can be used by educational authorities and policymakers to support student learning.

NOTE TO SELF : I WANT TO CHANGE THE FIRST SENTENCE IN ABSTRACT

1 Introduction

In 2020, the World's Health Organization (WHO) declared the coronavirus disease, commonly known as COVID-19, a public health emergency of international concern (World Health Organization 2020). As an airborne disease, COVID-19 was highly contagious from person to person, which made virtually all lifestyle activities a health risk during the pandemic. During this time, hosting in-person classroom activities and instructional periods were also considered a health hazard and were strongly discouraged. It attempts to support students and staff despite the uncertainty and unpredictability of the pandemic, school leaders and authorities implemented alternative learning models which offered students the opportunity to continue their studies in a safe way. In the United States, hybrid and virtual schooling modes were commonly adopted in 2020-2021 in response to the pandemic. **ANOTHER TRANSITION SENTENCE?**

*Code and data are available at: [LINK](#).

In this paper, I am interested in the impact of different schooling methods on US students' pass rates on state standardized exams in the pandemic school year in 2020-2021. Using district-level databases from 11 states and well as data from the National Center of Educational Statistics (NCES), I explore the changes in students' pass rates in 2020-2021 for more insights on the influence of the district's chosen learning model on the overall learning outcomes I find that, despite having the same curriculum, the decline in student pass rate is notably more obvious for schools who had a higher share in virtual or hybrid schooling modes. More specifically, (pass rates in English + Math) which results in an overall decline of _____% during the pandemic year. Also, I also find that schools with higher shares of Black and Hispanic students saw a greater decline in student pass rates. **MODIFY THIS IF NEEDED AFTER DISCUSSION** I can maybe show some stats here already??

The remainder of this paper is structured as follows. Section 2 discusses the data collection and the studied variables. Section 3 builds a model that suggests a relationship between learning models and student pass rates. Section 4 presents results and findings of the exploration for the dataset with the help of visualized data. Section 5 explores further insights from section 4 and discusses a few weaknesses and limitations of this study. This section also suggests potential next steps following this paper.

2 Data

The data was cleaned and processed using the statistical programming language R (R Core Team 2022). Statistical libraries such as `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2021), `arrow` (Richardson et al. 2023) and `here` (Müller 2020) are leveraged in the data processing as well.

2.1 Source and Data Collection

The paper and raw data used for replication is obtained from “Pandemic Schooling Modes and Student Test Scores: Evidence from US School Districts” (Jack et al. 2023), published in the American Economic Association's *American Economic Review: Insights* (AEA 2024). The downloaded data is built from district-level schooling mode data from the 2020-2021 schooling year and from district-level standardized assessment data from Spring 2019-2019 and 2021. I also downloaded additional district-level demographic data for this investigation. Detailed explanation of each data source and data collection method is given below.

2.1.1 District-Level Schooling Methods

Data on district-level schooling methods are downloaded from the COVID-19 School Data Hub (COVID-19 School Data Hub 2021). This is a public database which aggregates state-sourced

data to provide information on schooling modes and learning models by school districts during the 2020-2021 school year. Typically, the state-sourced data are State Education Agencies (SEA).

Table 1: First Six Rows of the Schooling Modes Data by District ID in 2020-2021 School Year

District ID	In-person Model (%)	% Virtual Model (%)	Hybrid Model (%)
807230	85.87	14.13	0.00
802430	85.87	6.01	8.13
807230	85.87	14.13	0.00
806750	21.55	6.01	72.44
802580	0.00	27.56	72.44
805550	93.99	6.01	0.00

For each district in the United States, Table 1 provides information on the percentage of the academic spent in each of type of schooling methods. To collect this data, the COVID-19 School Data Hub Team submitted data requests to state education agencies. They requested records of learning models used by schools and/or districts in the 2020-2021 academic school year. The renewal frequency of the data depends on on each school and districts where some would send new records weekly, while others would send the records in monthly. States who provided data monthly, bi-weekly or weekly during 2020-2021 year were included in this analysis.

2.1.2 District-Level Assessment Data

Data on district-level assessment results are collected from the Departments of Educations for the studied states. The respective departments collect data via surveys from constituent schools, school authorities and school boards (US Department of Education 2023). To evaluate changes in students' pass rates in math and ELA, data on test scores between Spring 2016-2019 and 2021 were extracted. Table 2 is a sample of this data and is organized by year of study between grade 3 and grade 8.

Table 2: First Six Rows of the Pass Rates by District ID, Subject and Grade from 2016-2019

District ID	Subject	Year	Grade 3 (%)	Grade 4 (%)	Grade 5 (%)	Grade 6 (%)	Grade 7 (%)	Grade 8 (%)
801950	math	2016	17.2	11.2	7.9	8.6	8.3	3.8
802430	ela	2016	28.9	46.3	41.2	27.1	27.8	26.5
802580	math	2016	37.3	26.1	31.1	21.7	21.7	13.7
805550	ela	2016	21.7	25.2	24.3	18.8	27.4	30.2
806750	math	2016	44.9	51.7	42.6	23.0	10.1	25.6

District ID	Subject	Year	Grade 3 (%)	Grade 4 (%)	Grade 5 (%)	Grade 6 (%)	Grade 7 (%)	Grade 8 (%)
806900	ela	2016	33.3	41.8	39.0	32.8	37.1	39.8

The downloaded dataset includes states which had at least two years of pre-pandemic test data available and had no significant changes to the assessment content. Certain states such as Alaska, Nevada and New York were excluded from the analysis as these states presented low assessment participation rate in 2021.

2.1.3 District-Level Demographic Data

Data on student demographic are downloaded from the National Center for Education Statistics (NCES), accessible via the Urban Institute’s Education Data Portal. Table 3 includes information on district-level share of enrolled students by race and ethnicity in the 2020-2021 school year. This information is collected through voluntary surveys by district.

Table 3: First Six Rows of Demographic Shares by District during the 2020-2021 School Year

District ID	State	Black Students (%)	Hispanic Students (%)	Other Students (%)
802430	CO	1.79	37.11	5.79
802580	CO	2.11	48.05	6.97
805550	CO	2.11	65.00	5.94
806750	CO	1.12	27.33	5.04
807230	CO	1.68	74.37	7.52
807230	CO	1.68	74.37	7.52

Based on the state selection criteria previously discussed, the final analysis data captures 11 majors states in the United States: Colorado (CO), Connecticut (CT), Massachusetts (MA), Minnesota (MN), Mississippi (MS), Ohio (OH), Rhode Island (RI), Virginia (VA), West Virginia (WV), Wisconsin (WI) and Wyoming (WY).

2.2 Variables of Interest

2.2.1 Schooling Modes

The possible schooling modes are in-person, virtual and hybrid learning models. In the analysis data, the schooling modes are defined and determined as follows:

- In-person: All or most students have access to traditional in-person instruction five days a week.

- Virtual: All or most students receive instruction online five days a week. Online instruction includes synchronous, asynchronous or a combination of synchronous and asynchronous activities.
- Hybrid: Schooling modes that do not correspond to any of the previous two models. Usually, this is a combination of the previous two.

2.2.2 Pass rates

Pass rates are calculated by dividing the number of students who passes standardized assessment by the number of students who took the exam in the district. For the purpose of this analysis, the pass rates for math and ELA assessments were considered. To count a pass, the student must score proficient or above on the selected state assessments.

3 Model

Through the exploration and the analysis of the data, I discover there is a correlation between learning model shares and students' pass rates on state assessments. To infer the efficiency of each learning models, we construct a Bayesian generalized linear model.

3.1 Model set-up

The model is run in R (R Core Team 2022) using the `rstanarm` package of Goodrich et al. (2022) using the default priors from `rstanarm`. The estimating equation is as follows:

$$Y_i = \beta_0 + \beta_1 \times \text{in-person} + \beta_2 \times \text{virtual} + \beta_3 \times \text{hybrid} + \epsilon \quad (1)$$

$$\beta_0 \sim \text{Normal}(0, 2) \quad (2)$$

$$\beta_1 \sim \text{Normal}(0, 2) \quad (3)$$

$$\beta_2 \sim \text{Normal}(0, 2) \quad (4)$$

$$\beta_3 \sim \text{Normal}(0, 2) \quad (5)$$

In this model, we define the variables as follows:

- Y_i is students' pass rates on the math and ELA state standardized assessments.
- β_0 is the coefficient for intercept.
- β_1 is the coefficient of the percentage of time spent in an in-person learning model.
- β_2 is the coefficient of the percentage of time spent in a virtual learning model.

- β_3 is the coefficient of the percentage of time spent in a hybrid learning model.
- ϵ is the fixed effects from districts, years and location.

3.2 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular... how to justify a model????

4 Results

results of model + summary() do the graphs with the dots (see what other people did)

5 Discussion

mental health component of virtual learning (social kakashi, health kakashi)

transition to virtual => learning to adapt a lot of trial and error

difference between grades? => younger had less difference since their curriculum is easier to understand?

difference between demographic => maybe there was something to discuss about resources

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

5.4.1 Weaknesses

A weakness in using data from the

weakness 1 : missing data => data failed test weakness 2 : definiton of in-person instruction. Access to in-perosn does not mean everyone actually did it online. Some may have opted to do it online for health reasons.

5.4.2 Next steps

testing are only one way to evaluate student learning such as peer assessment + self-assessments

References

- AEA. 2024. *The American Economic Association Portal*. *The American Economic Association*. <https://www.aeaweb.org/>.
- COVID-19 School Data Hub. 2021. *Percentage of School Year Spent in-Person, Hybrid, or Virtual, District Overall Shares*. <https://www.covidschooldatahub.com/data-resources>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Jack, Rebecca, Clare Halloran, James Okun, and Emily Oster. 2023. *Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts*. American Economic Review: Insights. <http://doi.org/10.3886/E168843V1>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- US Department of Education. 2023. “An Overview of the u.s. Department of Education.” <https://www2.ed.gov/about/overview/focus/what.html>.
- 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>.
- World Health Organization. 2020. *WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020*. World Health Organization. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.