The Impacts of COVID-induced School Closures on Inequality in Education Outcomes

Peter Ma[1] and Wang LingZhen[1#]

[1]Beijing No. 4 High School International Campus, China
[#]Advisor

ABSTRACT

I use online math participation and student math progress data from online learning platform Zearn to study the impacts of COVID-induced school closures on inequality in education outcomes using an event study approach. There were sharp declines in online math participation and progress following school closures but both outcomes were back to the pre-COVID levels in 2021 May. The pandemic widened socioeconomic inequalities in education outcomes transitorily, where most of the post-COVID socioeconomic gaps had vanished by 2021 May. Areas with lower income, more students in rural schools, worse computer and Internet access, and less higher education completion (measured by the percent of adults having a B.A.) saw more learning losses but only in the remaining 2020 spring semester.

1. Introduction

The outbreak of the COVID-19 pandemic in 2020 forced more than 160 countries to mandate school closures, which impacted at least 1.6 billion students (World Bank, 2020). As schools turned to fully online instruction, there have been concerns that the closures may hurt disadvantaged students more and widen education inequalities (Armitage & Nellums, 2020). Possible mechanisms include exacerbated food insecurity faced by poor students and socioeconomic background differences such as unequal access to digital technologies (Van Lancker & Parolin, 2020). Therefore, it is critical to characterize students' educational achievements under the influence of the COVID-19 pandemic, especially the gaps in learning outcomes between children from lower and higher socioeconomic backgrounds.

This paper uses real-time data from a widely used online learning platform Zearn to document the changes in online math participation and student progress as a result of the COVID-induced school closures. To characterize education inequality, I combine Zearn data with county- and state-level demographic characteristics including income, computer and Internet penetration, school rurality, and higher education completion. I use an event study approach to examine how differences in online math participation and progress between students from advantaged and disadvantaged areas vary week by week.

I document two main findings. First, there were rapid declines in online math participation and progress following school closures but both outcomes were back to the pre-COVID levels in 2021 May. In the remaining 2020 spring semester, online math participation and math progress decreased by 20 percentage points on average. This declining trend was transitory and decayed with time, where participation was almost back to the pre-COVID level and progress was even higher in May 2021. It suggests that students and schools got more and more adapted to remote learning as the pandemic developed.

Second, I show that the pandemic widened socioeconomic inequalities in education outcomes transitorily, where most of the post-COVID socioeconomic gaps had vanished by 2021 May. The gaps between counties of above and below nationwide median household income were about 15-30 percentage points for engagement and 40 percentage points for badges earned in the remaining 2020 spring semester, then shrank to levels close to zero. Similarly, areas with more students in rural schools, worse computer and Internet access, and less higher education completion (measured by the percent of adults having a B.A.) saw more learning losses but only in the remaining 2020 spring semester.

I also show that student outcomes from Zearn are positively correlated with Google search intensity of Zearn, which suggests that search intensity is a useful proxy for the corresponding learning resource. Student outcomes from Zearn are also positively correlated with search intensity of alternative platforms of Zearn, which suggests that students who suffer from larger learning losses in Zearn also use other online resources less. Therefore, the socioeconomic gaps in Zearn outcomes are unlikely to be due to the substitution from Zearn to alternatives.

This paper adds to three strands of literature. First, this work shows that educational achievement is sensitive to circumstances, including family background and neighborhood environment. Prior work on inequality finds that inequalities in early family environments and investments in children affect the development of capabilities, which affects adult attainments as foundations laid down earlier (Cunha & Heckman, 2009; Heckman, 2006). Students from families with low socioeconomic status suffer from greater learning losses over summer vacation in contrast with those who have a more advantaged family background (Alexander et al., 2007; Cooper et al., 1996). Apart from family background, better neighborhoods also have significant positive effects on future outcomes such as college attendance rates and earnings through childhood exposure (Chetty & Hendren, 2018a, 2018b). My results suggest that the effects of school closures on education outcomes differ by socioeconomic background, even though there were no systematic pre-COVID differences.

Second, this work provides new evidence on the digital divide between students in socioeconomically advantaged and disadvantaged areas. Previous literature documents pre-COVID socioeconomic gaps in three layers of digital divide: access to computer and the Internet, access of skills and usage, and outcomes of technology use (Dijk & Hacker, 2003; DiMaggio & Hargittai, 2001; Vigdor et al., 2014). Van de Werfhorst et al. (2020) find that digital preparedness of students before the start of the COVID-19 pandemic varies by socioeconomic background. Explicitly touching on the first and third layers of digital divide, I complete this evidence by showing that the COVID-19 pandemic compounded digital divide between socioeconomically advantaged and disadvantages areas.

Third, my work contributes to studies investigating the effects of COVID-19 induced school closures on student outcomes such as learning behaviors, time allocation, and standardized tests. Cai (2020) observe that exam scores of students from rural households are more negatively affected by school closures. Similarly, using standardized test outcomes in Netherlands and Belgium respectively, Engzell et al. (2021) as well as Maldonado and De Witte (2021) find significant learning losses and larger negative effects for disadvantaged students as a result of 2020 school closures. Bacher-Hicks et al. (2021) document gaps in search intensity of online learning resources between socioeconomically advantaged and disadvantaged areas. Andrew et al. (2020) find inequalities in time spent learning during lockdown between poorer and richer students using UK survey data. I provide the first work evaluating the effects of COVID-19 induced school closures over a longer time period instead of merely instantaneous responses. By using data from January 2019 through May 2021, I document that the learning gaps correlated with differential socioeconomic status decayed over time.

My findings show that it is critical for policymakers to target students from socioeconomically disadvantaged areas when making policies intending to recover the COVID-induced learning losses. Although both student participation and progress have been back to the pre-COVID levels and socioeconomic inequalities have mostly vanished, there can still be long run effects as a result of temporary education outcome gaps.

One limitation of this paper is that the underlying mechanism of socioeconomic gaps in education outcomes is not fully studied. In particular, in this remote instruction setting, there might be some new drivers of education inequality related to online learning, such as different skills and usage of online learning resources. To reduce education inequalities associated with socioeconomic status, future research may focus on exploring the mechanism behind them.

2. Data and Empirical Strategy

2.1 Data

The ideal data set for studying the impact of COVID-19 on education inequality might be student-level administrative data collected both before and after the outbreak of pandemic, but it is hardly possible to find one. Alternatively, I use data about education outcomes from a widely used nonprofit curriculum publisher Zearn, which is made publicly available at the Opportunity Insights Economic Tracker (Chetty et al., 2020). Zearn Math, published by Zearn, is a K-5 math program combining both hands-on instruction and software-based digital learning. In 2020, Zearn Math was used by 1 in 4 elementary school students nationwide. Many schools had integrated Zearn Math as part of their math curriculum before the pandemic and continued to use it in remote instruction during the pandemic. These facts suggest that data from Zearn is representative of online learning outcomes for students in the U.S. as a whole.

I use online math participation and student math progress as outcome variables. Online math participation, referred to as "engagement" on Zearn, measures the percent change in number of students using Zearn Math each week relative to the January 6-February 2, 2020 indexing period. Student math progress, referred to as "badges earned" on Zearn, measures the percent change in number of lessons completed by students each week on Zearn Math relative to the January 6-February 2, 2020 indexing period. Both variables are available only among schools that already used Zearn Math in course instruction before the pandemic. Because both online math participation and student math progress indices during summer and winter breaks are volatile and do not reflect the true values, observations during major breaks are excluded. I use county-level data from January 2019 to May 2021 in main analysis and state-level data for the same period as supplement. Table 1 presents summary statistics for engagement and badges earned in panel A. Engagement falls into the range from -100 to 150 because Chetty et al. (2020) winsorize values exceeding 150. Badges earned ranges from -100 to 300 due to a similar winsorizing process. The mean of engagement is -14.68, which means full sample average engagement is 14.78% lower than the indexing period (January 6-February 2, 2020) average. Full sample average badges earned is 3.99% lower than the indexing period average.

Although Zearn is widely used among elementary schools in the U.S., there are a lot of alternatives to Zearn Math. Given the fixed cost of integrating a learning platform into curriculum, it is unlikely that schools or parents that had been using Zearn before March 2020 turned to other online resources after the outbreak of the COVID-19 pandemic. However, to rule out this possibility, I use search intensity of Zearn as well as alternatives to Zearn from Google Trends as proxies for the popularity of each platform. The search intensity provided by Google Trends is a measure of the popularity of the specified term as a fraction of total searches in that location. Values are normalized to a scale from 0 to 100 and are calculated weekly. Because Google does not provide county-level search intensity data, I use state-level search intensity from January 2019 to May 2021 instead. The four most popular keywords of alternatives to Zearn I identify are "Reflex Math", "IXL Math", "Freckle Math", and "iReady Math", which are branded online math learning resources for K-12 students. I sum together the search intensity of the four keywords as a single measure of the popularity of major alternatives to Zearn Math. During January 2019-May 2021, the average search intensity of keyword "Zearn" accounts for 19.78% of total searches, while alternatives to Zearn accounts for 34.5%.

To characterize education inequality, I use pre-COVID county- and state-level demographic data from the 2019 American Community Survey (ACS) and the 2018 Stanford Education Data Archive (SEDA). I use the following measures from ACS: percent of households with broadband Internet, percent of households with a computer, and percent of adults with a B.A. As of 2019, 72.5% households had a computer and 77.97% households had assess to broadband Internet on average, while 23.83% adults held a bachelor's degree or above on average. SEDA provides additional information, including the percent of students in rural areas and the logarithm of median family income. As of 2018, the mean of the fraction of students in rural schools was 41.45% and 58% of the counties had local median family income above the national median.

Table 1
Summary Statistics.

	Mean	SD	Min	Max
(A) Education outcome				
Engagement (Online math participation)	-14.78	38.15	-100	150
Badges (Student math progress)	-3.99	52.77	-100	300
(B) County socioeconomic traits				
Median family income above nationwide median	0.58	0.49	0	1
Log of median family income	10.77	0.25	9.92	11.61
Percent students in rural local schools	41.45	30	0	100
Percent households with a computer	72.5	9.46	33.65	95.05
Percent households with broadband internet	77.97	8.32	34.77	96.03
Percent adults with a B.A. and above	23.83	10.4	7.43	74.46
Observations	114938			
(C) Search intensity from Google Trends				
Search intensity of Zearn	19.78	22.17	0	100
Search Intensity of Alternatives to Zearn	34.5	32.77	0	224
Relative Change of Search Intensity of Zearn	-10.33	133.07	-100	1367
Relative Change of Search Intensity of Alternatives to Zearn	-22.34	79.05	-100	800
Observations	4029			

Notes: Panel A presents summary statistics of data from Zearn, an online math learning platform. Panel B presents county socio-economic traits from the 2019 American Community Survey (ACS) and the 2018 Stanford Education Data Archive (SEDA). Data in Panel A and B are county-level measures. Panel C presents summary statistics of state-level data from Google Trends.

2.2 Empirical Strategy

I first estimate changes in education outcomes as a result of school closures forced by the COVID-19 pandemic, using a week-to-week event study specification. The event study specification is :

$$y_{it} = \alpha + \sum_{j=1}^{9} \beta_j lagj_t + \sum_{j=1}^{59} \gamma_j leadj_t + \mu_i + \delta_{w(t)} + \lambda_{y(t)} + u_{it}, \tag{1}$$

where i denotes county and t denotes calendar time. The outcome variables are online math participation (engagement) and student math progress (badges earned). $lagj$ and $leadj$ are indicators for week t falling before or after March 15, 2020 by j weeks. For example, $lag1$ indicates one week prior to March 15, 2020, which is March 8, 2020, and $lead1$ indicates one week after March 15, 2020, which is March 22, 2020. Most states ordered or recommended school closure starting from mid-March and hence I choose March 15, 2020 as the omitted baseline category ($j = 0$). I exploit the panel nature of the data by including county, week of year $w(t)$, and year $y(t)$ fixed effects. The vector of weekly indicators covers all weeks from January 2020 to May 2021, leaving the 2019 data to identify week of year effects. Thus, the coefficient β_j (γ_j) measures the deviation of education outcome j weeks before (after) March 15, 2020 from calendar predicted outcome relative to that baseline week.

To study how education outcomes change differentially by socioeconomic traits, I modify Eq. (1) by adding interactions between the pre- and post- COVID indicators and socioeconomic traits. The modified event study specification is :

$$y_{it} = \alpha + \sum_{j=1}^{9} \beta_j lagj_t \times Trait_i + \sum_{j=1}^{59} \gamma_j leadj_t \times Trait_i + \sum_{j=1}^{9} \sigma_j lagj_t + \sum_{j=1}^{59} \tau_j leadj_t + \mu_i + \delta_{w(t)} + \lambda_{y(t)} + u_{it}. \quad (2)$$

Here, $Trait$ represents the indicator for whether a county is of above or below nationwide median household income, the logarithm of median family income, percent of schools in rural areas, percent of households with broadband Internet, percent of households with a computer, and percent of adults with a B.A. in the county. The coefficients β_j and γ_j estimate differences in weekly math learning outcomes by socioeconomic traits.

Given the large number of coefficients in event study specifications, I also estimate the following first difference equation using the post-COVID sample:

$$y_{it} = \beta_0 + \beta_1 Trait_i + \delta_{w(t)} + \lambda_{y(t)} + u_{it}, \quad (3)$$

where t is restricted to the period beginning on March 23, 2020. Thus β_1 is interpreted as the post-COVID difference in education outcomes by county-level socioeconomic characteristics. Because both student engagement and badges earned are relative measures which are normalized to the indexing period January 6-February 2, 2020, it is challenging to explain the coefficients of the interaction term in the usual difference-in-differences specification associated with an event study specification. Therefore, to facilitate the interpretation of coefficients, I use a first difference specification instead of the commonly used difference-in-differences specification. All standard errors are clustered at the county level.

3. Results

3.1 Effects on Education Outcomes

I present the event study estimates of Eq. (1) in Figure 1. The figure on the left shows the coefficient estimates where the outcome variable is student engagement. The normalized values of engagement are not significantly different from or close to baseline engagement prior to March 15, 2020. Since engagement is a relative measure, which is normalized towards the indexing period January 6-February 2, 2020, panel A shows that student online math participation was quite stable during the first ten weeks in 2020. There was a significant decline in math participation right after March 15, 2020, where the magnitude was about 40 percentage points in the first two weeks and decreased to about 20 percentage points at the end of the 2020 spring semester. This declining trend was transitory and decayed with time, where in 2020 fall semester student engagement was about 10 percentage points less than baseline engagement. 2021 spring semester saw a more volatile pattern, but engagement was almost back to pre-COVID level in May. Panel B of Figure 1 shows a very similar pattern of changes in badges earned (lessons completed) before and after March 15, 2020. After plummeting by about 20 percentage points in the following four weeks, student math progress was back to pre-COVID level and went even higher in 2020 fall and 2021 spring semester.

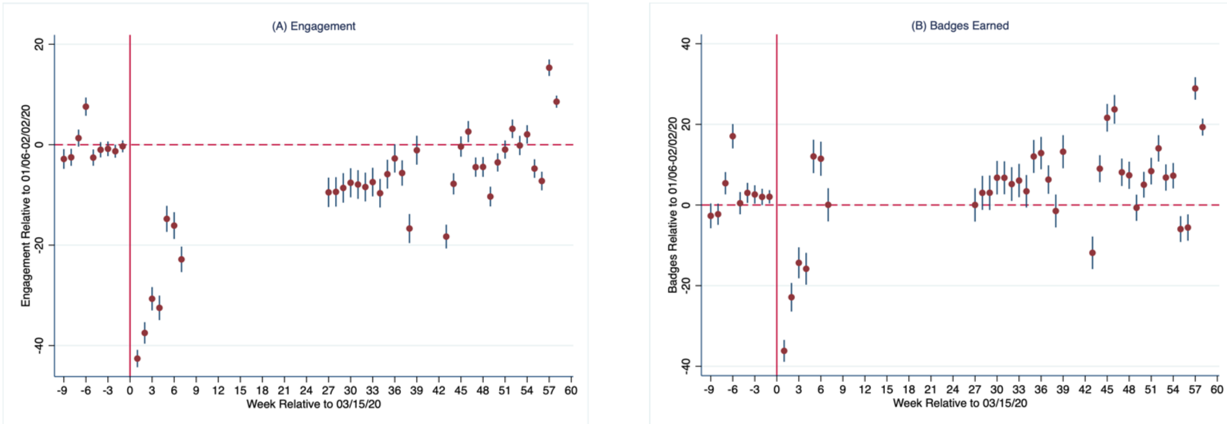


Figure 1. Event Study of Education Outcomes.

Notes: The figure above shows event study coefficients estimating changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

Event study analyses show that student engagement and badges earned decreased less in areas with higher income after the outbreak of the COVID-19 pandemic. Figure 2 presents the regression-based weekly difference in education outcomes from Eq. (2) where the vector of weekly indicators is interacted with whether the county is of above nation-wide median family income, controlling for county, week of year, and year fixed effects. There is no significant difference in engagement or badges earned between high- and low- income counties before school closures. However, high-income counties saw smaller declines in both engagement and badges earned than low-income counties after March 15, 2020. The gaps were about 15-30 percentage points for engagement and 40 percentage points for badges earned in the remaining 2020 spring semester, then shrank to levels close to zero for both measures but did not vanish completely. Figure 3 shows the coefficient estimates of difference in education outcomes from Eq. (2) where the vector of weekly indicators is interacted with continuous local median family income instead of the binary indicator. In 2020 spring semester, one percent higher local median income is associated with over 0.6 percentage points more of student engagement and 1 percentage point more of badges earned.

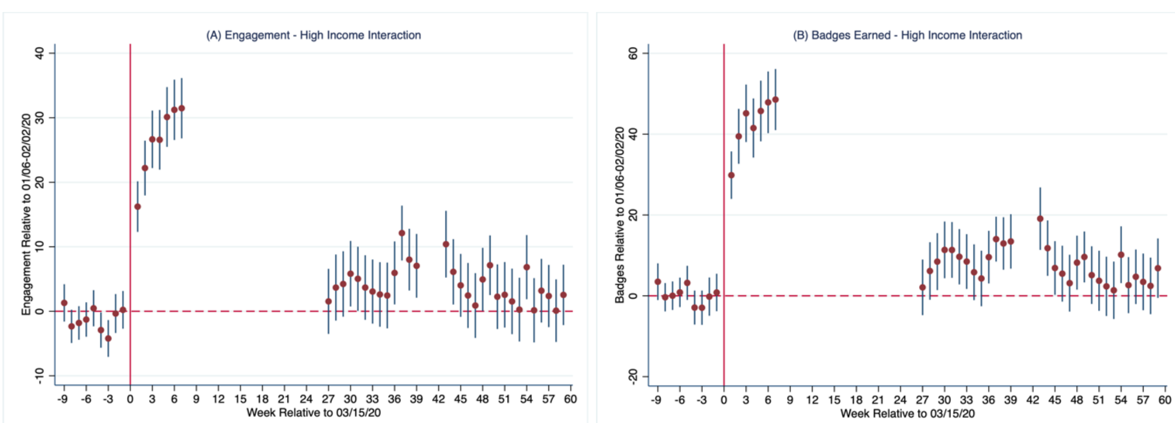


Figure 2. Event Study of Education Outcome Gap by Income Level.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 between counties of above and below nationwide median family income. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

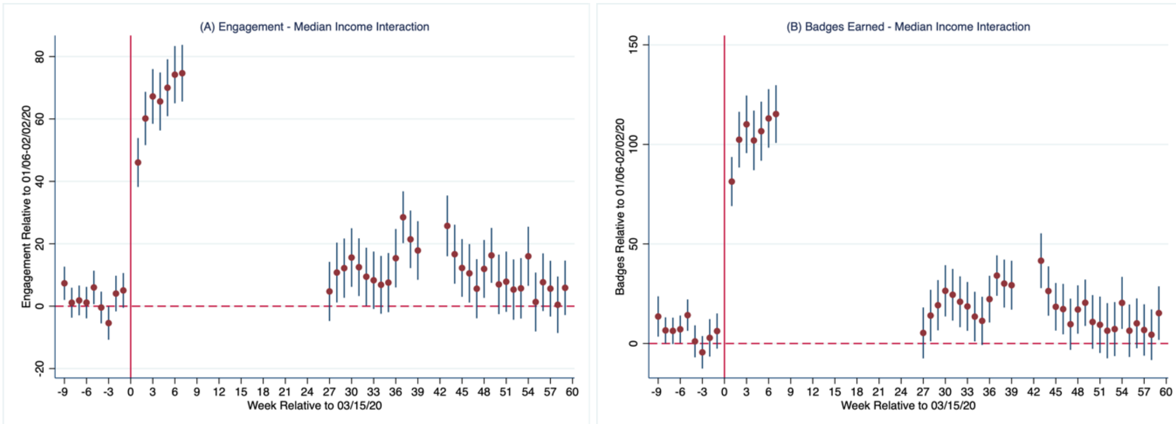


Figure 3. Event Study of Education Outcome Gap by the Logarithm of Local Median Family Income.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 by the logarithm of local median family income. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

The pandemic also widened gaps in education outcomes by school rurality, technological access, and higher education completion, but in a more transitory manner. Figure 4 shows that one more percentage point of students in rural schools decreased engagement by about 0.3 percentage points and badges earned by about 0.5 percentage points in 2020 spring semester.

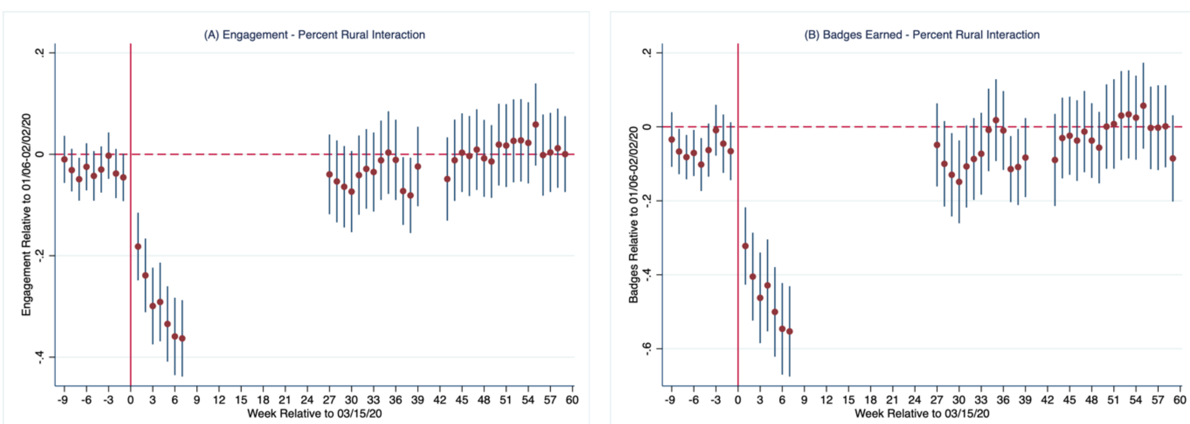


Figure 4. Event Study of Education Outcome Gap by Percent of Students in Rural Schools.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 by the percent of students in rural schools. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 1 (2022)

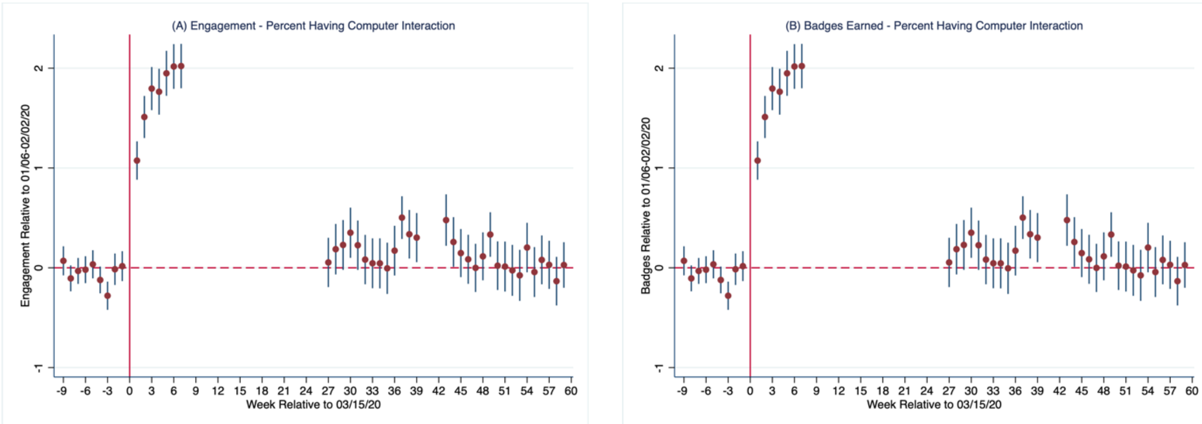


Figure 5. Event Study of Education Outcome Gap by Percent of Households Having Computer.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 by the percent of households having computer. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

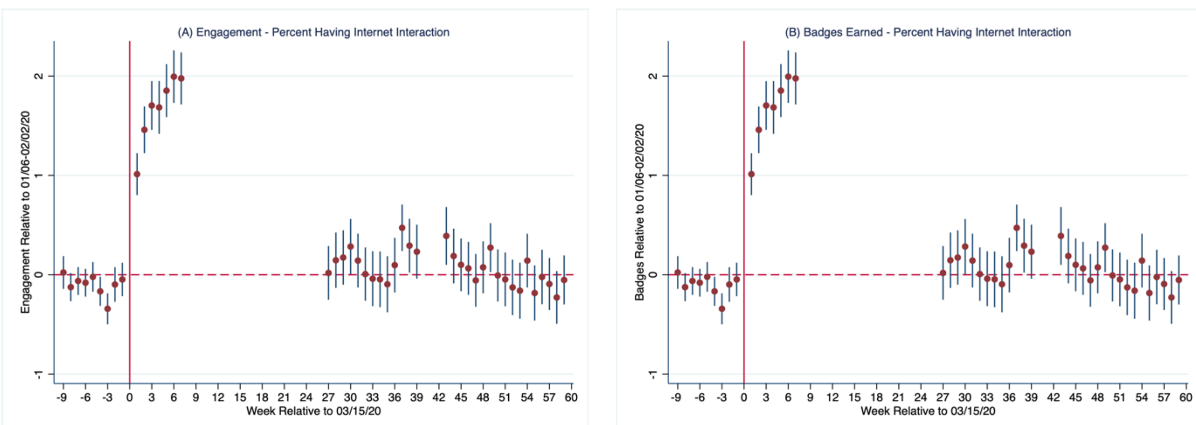


Figure 6. Event Study of Education Outcome Gap by Percent of Households Having Broadband Internet.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 by the percent of households having broadband Internet. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

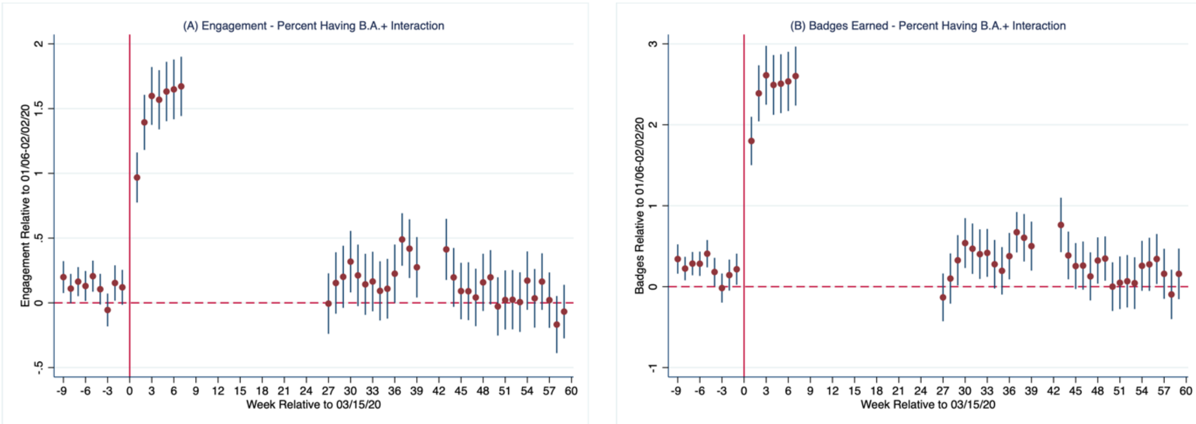


Figure 7. Event Study of Education Outcome Gap by Percent of Adults with a B.A.

Notes: The figure above shows event study coefficients estimating the differences in changes in online math participation and student math progress from math learning platform Zearn compared to the baseline week March 15, 2020 by the percent of adults with a B.A. The regressions include fixed effects for county, week of year, and year. The lines denote the 95% confidence interval and standard errors are clustered at the county level. The sample contains data from January 2019 to May 2021, excluding summer and winter breaks.

After the school closures, but this gap faded out quickly and was neither significant in 2020 fall nor 2021 spring semester. Figure 5 and 6 show that one percentage point increase in access to computer or broadband Internet increased online math participation and student math progress by approximately 1.5 percentage points, while the effect was only present in the several weeks following school closures. Figure 7 shows that one percentage point increase in people holding a bachelor's degree increased engagement by approximately 1.5 percentage points and badges earned by approximately 2.5 percentage points in 2020 spring semester, and then decayed and eventually vanished in 2021 spring semester.

Table 2

Post-COVID Differences in Education Outcomes by Socioeconomic Traits.

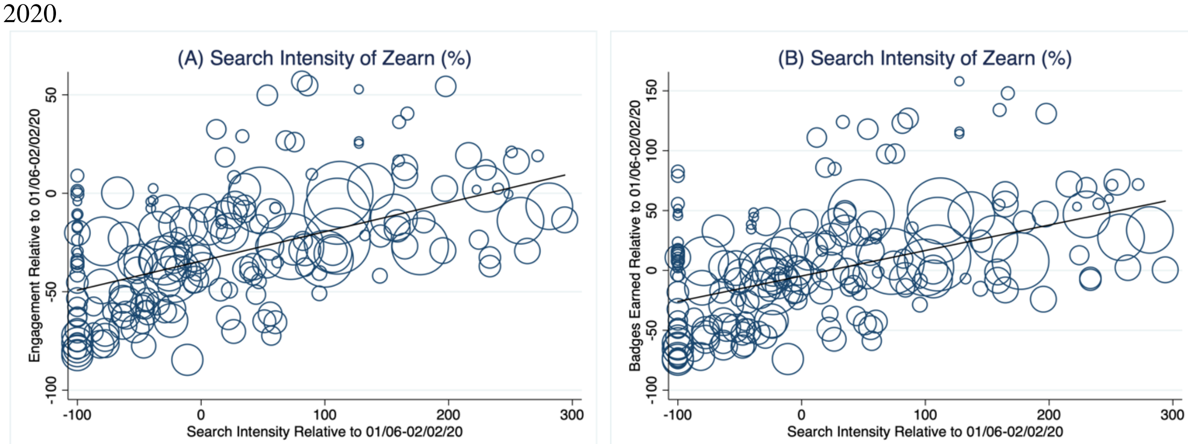
Dependent variable	Engagement	Badges Earned
	(1)	(2)
High income	9.085^{***}	12.71^{***}
	(1.864)	(2.623)
Log (median income)	17.20^{***}	22.74^{***}
	(3.633)	(5.085)
Percent students in rural schools	-0.0347	-0.0522
	(0.0304)	(0.0419)
Percent households with a computer	0.427^{***}	0.721^{***}
	(0.0908)	(0.126)
Percent households with broadband internet	0.389^{***}	0.652^{***}
	(0.0998)	(0.141)
Percent adults having a B.A.	0.210^{**}	0.341^{***}
	(0.0906)	(0.122)
Observations	53474	53474

Notes: Each coefficient comes from a separate regression that includes the independent variable in that row, week of year, and year fixed effects. The outcome variables corresponding to Columns 1 and 2 are normalized student online math participation and student math progress from Zearn respectively. The sample contains data from March 22, 2020 through May 2021, excluding summer and winter breaks. All standard errors are clustered at the state level. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

First difference analyses show that the average post-COVID difference in student math progress is larger than online math participation between areas with high and low socioeconomic status. Table 2 presents the estimates of post-COVID education outcome gaps based on Eq. (3), where column 1 and 2 show gaps in engagement and badges earned respectively. Restricting to the subsample starting from March 22, 2020, the estimates can be interpreted as the average post-COVID learning gaps. Comparing estimates in column 1 and 2 suggests that the magnitude of gaps in badges earned is noticeably larger than those in engagement by any socioeconomic dimension. For example, low-income areas experience a decrease in student engagement 9 percentage points more than high-income areas on average during the full post-COVID period, while the gap is a greater 13 percentage points in badges earned.

3.2 Are Students Substituting Zearn with Alternatives?

I strengthen the prior results by excluding the possibility that areas with high socioeconomic status substituted from Zearn to alternative online learning resources after school closures and therefore caused the gap in Zearn outcomes. First, I illustrate that search intensity is a useful proxy for education outcomes. Figure 8 presents the correlation between relative changes in state-level search intensity of keyword "Zearn" and education outcomes from Zearn during April 1-May 31, 2020, the period when both measures change intensely.[1] Both correlation coefficients and fitted lines are computed by weighting each state by its population. To match the way engagement and badges earned are normalized, I calculate the percent change of search intensity relative to the same indexing period, January 6-February 2, 2020.



Correlation = 0.57 Correlation = 0.54

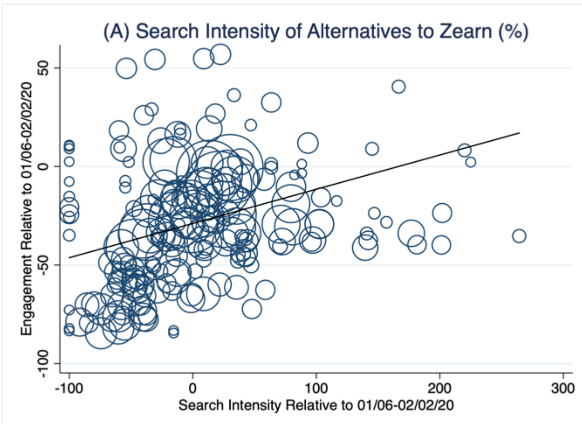
Figure 8. Correlation between Relative Changes in Search Intensity of Zearn and Education Outcomes.

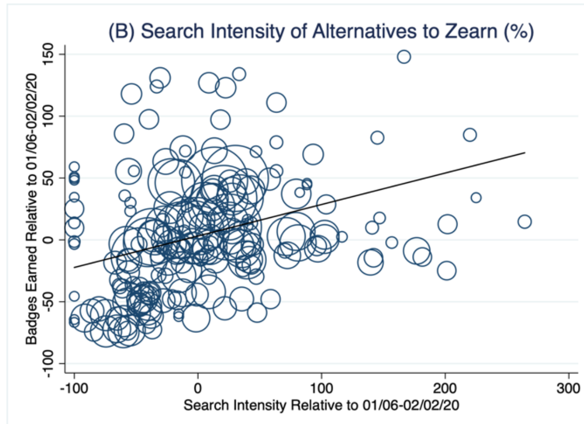
Notes: The figures above show the correlation between state-level changes in search intensity of keyword "Zearn" and online math engagement as well as student math progress from online learning platform Zearn. The data cover April 1-May 31, 2020. Both measures are normalized as percent changes relative to January 6-February 2, 2020. Both correlation coefficients and fitted lines are computed by weighting each state by its population.

[1] The correlation coefficients between education outcomes and search intensity over the whole period, January 2019-May 2021, are 0.11 and 0.13 respectively, which are too weak for search intensity to serve as proxies for education outcomes. Therefore, I focus on the correlation only during April 1-May 31, 2020 in the supplementary analysis where search intensity data are used.

HIGH SCHOOL EDITION
Journal of Student Research

I also exclude the values of relative change in search intensity that exceed 300%. Figure 8 presents a positive correlation between education outcomes and search intensity of the corresponding learning platform. The correlation coefficient between search intensity of "Zearn" and student engagement is 0.57 and that for badges earned is 0.54, which suggests that search intensity of a learning platform is a plausible proxy for students' education behaviors and outcomes related to that platform.





Correlation = 0.34 Correlation = 0.33

Figure 9. Correlation between Relative Changes in Search Intensity of Alternatives to Zearn and Education Outcomes.

Notes: The figures above show the correlation between state-level changes in search intensity of alternatives to Zearn and online math engagement as well as student math progress from online learning platform Zearn. Search intensity of alternatives of Zearn sums together search intensity of four keywords: "Reflex math", "IXL math", "Freckle math", and "iReady math". The data cover April 1-May 31, 2020. Both measures are normalized as percent changes relative to January 6-February 2, 2020. Both correlation coefficients and fitted lines are computed by weighting each state by its population.

Second, Figure 9 shows that education outcomes from Zearn are also positively correlated with search intensity of alternative learning platforms to Zearn, though to a lesser extent. The search intensity of alternatives to Zearn is also computed as percent change relative to January 6-February 2, 2020. The correlation coefficients between search intensity of alternatives to Zearn and Zearn outcomes are 0.34 and 0.33 for engagement and badges earned during April 1-May 31, 2020 respectively. Students' achievement on alternative platforms moves in the same direction as that on Zearn, which suggests it is unlikely that students simply use Zearn less and turn to its competitors after the outbreak of the pandemic.

Third, by regressing search intensity on post-COVID and/or high-income indicators, I find no evidence on the explanation that students from high-income areas substitute from alternative online resources to Zearn. Table 3 shows the regression-based difference in search intensity between states with high and low income. Columns 1 and 2 present the difference-in-differences estimates of the post-COVID gap where the outcome variables are raw search intensity values of Zearn and alternatives to Zearn respectively. I do not detect any significant post-COVID difference in searching behaviors regarding alternatives to Zearn between high- and low-income states. However, high-income areas have significantly higher post-COVID search intensity of Zearn, which coincides with the previous finding that student engagement and badges earned decrease less in high-income areas after the break of COVID pandemic. For comparison purposes, I regress the relative change of search intensity on high-income indicator using only the post-COVID subsample, which is parallel to the first difference specification in main analysis. Columns 3 and 4 present the first difference estimates in relative search intensity of Zearn and alternatives to Zearn respectively. The post-COVID difference in search intensity remains insignificant for alternatives to Zearn and that for Zearn is not significant either, which is probably because of the smaller sample size. In sum, there is no evidence on substitution between Zearn and other learning resources.

Table 3
Differences in Search Intensity by Income Level.

Dependent Variable	Diff-in-diff		First Difference	
	Zearn	Alternatives	Zearn	Alternatives
	(1)	(2)	(3)	(4)
Post-COVID * High income	8.316***	0.631		
	(2.843)	(4.804)		
Post-COVID	-14.23***	-24.18***		
	(2.493)	(3.447)		
High income			11.16	-2.647
			(7.258)	(4.364)
Observations	3022	3137	1365	1416
R^2	0.458	0.474	0.292	0.358

Notes: The outcome variables corresponding to Columns 1 and 2 are raw values of state-level search intensity of Zearn and alternatives to Zearn respectively. Search intensity is from Google Trends. Both regressions include state, week of year, and year fixed effects and use data from January 2019 to May 2021. The outcome variables corresponding to Columns 3 and 4 are changes of state-level search intensity of Zearn and alternatives to Zearn respectively, relative to the indexing period January 6-February 2, 2020. Both regressions include week of year and year fixed effects and use data from March 22, 2020 through May 31, 2021. Observations where relative search intensity changes more than 300% are excluded. All standard errors are clustered at the state level. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

4. Conclusion and Discussion

The main purpose of this paper is to study the impacts of COVID-induced school closures on inequality in education outcomes. I document a sharp but transitory decrease in online math participation and student math progress induced by the COVID-19 pandemic in the first seven weeks after March 15, 2020. Students' online math participation decreased by 20-40 percentage points and math progress decreased by 20 percentage points in 2020 spring semester. The declining trends decayed over time, where both education outcomes were back to pre-COVID level in 2021 spring semester. Though areas with both high and low socioeconomic status experience losses in education outcomes, the loss was substantially smaller in areas with high socioeconomic status. Specifically, areas with higher income level, higher availability of computer and broadband Internet, lower rurality, and higher fraction of highly educated residents had smaller post-COVID decreases in education outcomes. The COVID-induced education outcome gaps shrank over time and most of them eventually closed up in 2021 spring semester.

My findings show that it is critical for policymakers to target students from socioeconomically disadvantaged areas when making policies intending to recover the COVID-induced learning losses. Although both student participation and progress have been back to the pre-COVID levels and socioeconomic inequalities have mostly vanished, there can still be long run effects as a result of temporary education outcome gaps. Policymakers may focus on how to make sure schools have sufficient capability on distance teaching provision and effective involvement in home learning, how to make sure students have essential electronic devices for learning in the disadvantaged regions, and how to provide additional support to low-achieving students when they come back at school.

One limitation of this paper is that the underlying mechanism of socioeconomic gaps in education outcomes is not fully studied. In particular, in this remote instruction setting, there might be some new drivers of education inequality related to online learning, such as different skills and usage of online learning resources. To reduce education inequalities associated with socioeconomic status, future research may focus on exploring the mechanism behind them.

5. References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting Consequences of the Summer Learning Gap. *American Sociological Review*, *72*(2), 167-180. https://doi.org/10.1177/000312240707200202

Andrew, A., Cattan, S., Dias, M. C., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). Inequalities in Children's Experiences of Home Learning during the COVID-19 Lockdown in England*. *Fiscal Studies*, *41*(3), 653-683. https://doi.org/10.1111/1475-5890.12240 PMID - 33362314

Armitage, R., & Nellums, L. B. (2020). Considering inequalities in the school closure response to COVID-19. *The Lancet Global Health*, *8*(5). https://doi.org/10.1016/s2214-109x(20)30116-9

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, *193*. https://doi.org/10.1016/j.jpubeco.2020.104345

Cai, X. F., Jingcheng; Lu, Yi; Tang, Xiangming; Zhong, Songfa. (2020). *Assessing Inequality in the School Closure Response to COVID-19*.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & Team, t. O. I. (2020). *The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data*.

Chetty, R., & Hendren, N. (2018a). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, *133*(3), 1107-1162. https://doi.org/10.1093/qje/qjy007

Chetty, R., & Hendren, N. (2018b). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics*, *133*(3), 1163-1228. https://doi.org/10.1093/qje/qjy006

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research*, *66*(3), 227-268. https://doi.org/10.3102/00346543066003227

Cunha, F., & Heckman, J. J. (2009). THE ECONOMICS AND PSYCHOLOGY OF INEQUALITY AND HUMAN DEVELOPMENT. *Journal of the European Economic Association*, *7*(2-3), 320-364. https://doi.org/10.1162/jeea.2009.7.2-3.320 PMID - 20209045

Dijk, J. v., & Hacker, K. (2003). The Digital Divide as a Complex and Dynamic Phenomenon. *The Information Society*, *19*(4), 315-326. https://doi.org/10.1080/01972240309487

DiMaggio, P., & Hargittai, E. (2001). *From the 'Digital Divide' to 'Digital Inequality'- Studying Internet Use as Penetration Increases*.

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proc Natl Acad Sci U S A*, *118*(17). https://doi.org/10.1073/pnas.2022376118

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900-1902. https://doi.org/10.1126/science.1128898 PMID - 16809525

Maldonado, J. E., & De Witte, K. (2021). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*. https://doi.org/10.1002/berj.3754

Van de Werfhorst, H. G., Kessenich, E., & Geven, S. (2020). *The Digital Divide inOnline Education.Inequality in Digital Preparedness of Students and Schoolsbefore the Start of the COVID-19 Pandemic*.

Van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: a social crisis in the making. *The Lancet Public Health*, *5*(5), e243-e244. https://doi.org/10.1016/s2468-2667(20)30084-0

Vigdor, J. L., Ladd, H. F., & Martinez, E. (2014). Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry*, *52*(3), 1103-1119. https://doi.org/10.1111/ecin.12089

World Bank. (2020). *Guidance Note on Education Systems' Response to COVID19*.