Online Learning and the Education Gap: A Digital Footprint Approach

Yongheng Deng [1] Xinfei Huang [2] Teng Li [2] Shang-Jin Wei [3]

[1]University of Wisconsin-Madison

[2]Sun Yat-sen University

[3]Columbia University

May 23, 2023

Motivation: Online learning

- Unequal quality of education is a contributing factor to income inequality in both developed and developing countries
- US: tax reforms
- China: "common prosperity"
- Online learning has become more prevalent
  - Technology: smartphones and tablets; Apps; short videos
  - Shocks - pandemic; extreme weather events
- Effects of online learning on the education gap is unclear

Debate on online learning

No consensus on the effect of online learning on learning outcomes

- Possible negative effect? Opinion pieces in the *Economist*: the school closure due to the pandemic very bad for student learning.
  - "How Covid-19 Is Interrupting Children's Education – Almost A Billion Children Have Seen Their Schools Close" (*Economist*, 03/19/2020).
  - Halloran, Jack, Okun, and Oster (2023): across US school districts in 2020-2021, online and hybrid teaching associated with lower testing scores. However, times on online and hybrid learning correlated with severity of Covid outbreak
- Possible beneficial effects?
  - A meta analysis of 21 RCT research publications and 6 non-RCT analysis show **medical undergraduates**' learning outcomes are generally better for the treatment (online learning) group than the control (offline) group (Gao et al. 2022).
  - mechanism: Increased flexibility and comfort of the home environment
  - younger children could be different
  - no study on differential effects of rich vs poor students

Open Research Question

- How does the school closure affect the gap in learning outcomes?
- Role of parents in the gap (to do)

This paper: How prolonged online learning, triggered by pandemic-related school closures, affects the educational gap between rich and poor children.

This Paper

We explore a digital footprint approach to overcome a lack of systematic data on identifying the subjects of interest (students), their family economic status, and learning outcomes

- **Data:** Universe of identity-masked and geocoded cell phone usage information from a major Chinese telecommunication service provider in Guangdong.
- **Subjects of interest:** Families with a child who is a primary school student before the summer and enters a middle school in the fall semester.
- **Learning Outcome:** Admission to a top middle school (based on test scores)
- **Family wealth:** Housing prices of residence and their shopping behavior.

Empirical Strategy: to link educational outcome to family wealth and length of school closure (online learning)

$$E_{ict} = \alpha_1 W_{i,t-s} + \alpha_2 W_{i,t-s} \times SC_{ict} + \alpha_3 SC_{ict} + \delta_c + \eta_t + \epsilon_{ict} \quad (1)$$

E_{ict} = learning outcome for student i in city c & year t.

$W_{i,t-s}$ = family wealth for student i.

SC_{ict} = duration of school closure (online learning) in city c and year t.

δ_c : city fixed-effect.

η_t : year fixed-effect.

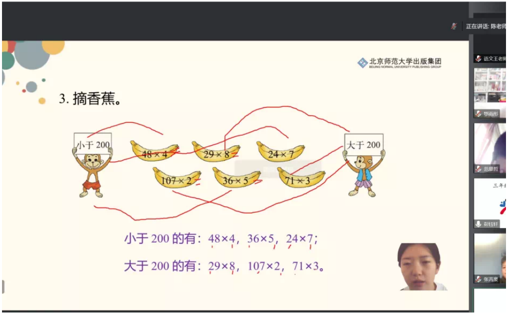
This Paper

We find:

- There is an educational gap between the rich and poor children even without school closure - children in the top quartile wealth bin are 12.8 percentage points more likely to enter a top middle school than those in the bottom quartile bin
- School closure /online learning has widened the educational gap -in places where schools are closed for half a year, children in the top quartile wealth bin are now 23.6 percentage points more likely to enter a top school than those in the bottom quartile.
- Differences in supplementary app usage and parental accompaniment are possible contributing factors.

Motivation: Online learning

- In response to the start of Covid in late January 2020, Chinese schools switched to online teaching (mandated by the Ministry of Education)
- This marks the first time that online teaching is officially used on a large scale



Background: School closure and spatial variations

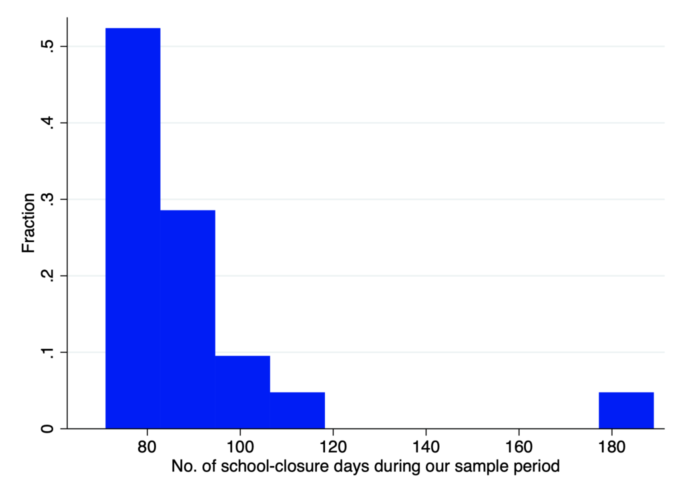
- Under the zero-Covid policy, local governments order school closures when there is a Covid outbreak
- Closure even with a relatively small number of cases. Neither children nor parents are likely to be sick.
- Online teaching on school closure days.
- A typical primary school student experienced 88 days of school closure, with a standard deviation of 41 days.
- Outside school hours, students may take commercial online classes on weekends or evenings. True both before and during Covid (until 2021)

猿辅导

学而思网校
Xueersi Online

Background: School-closure days

This graph presents the histogram of school closure days for each city during our sample period.

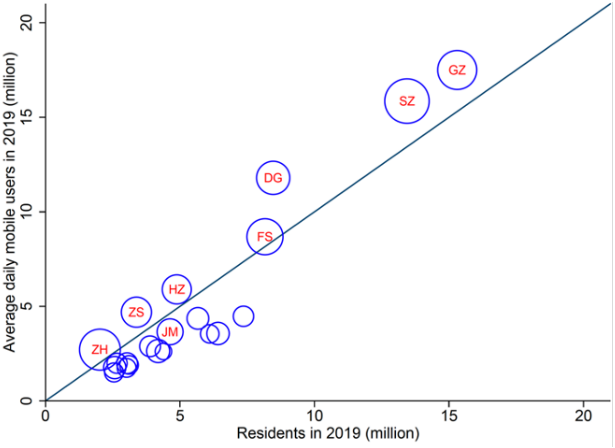


Data

Cities	Population 2019 (million)	GDP 2019 ($ billion)	No. of key middle schools	No. of non-key middle schools	School closure days in 2020	School closur days in 2021
Shenzhen (SZ)	13.44	390.25	21	105	71	0
Guangzhou (GZ)	15.31	342.44	42	133	78	31
Foshan (FS)	8.16	155.81	10	50	71	26
Dongguan (DG)	8.46	137.43	6	26	78	9
Huizhou (HZ)	4.88	60.54	9	13	71	23
Zhuhai (ZH)	2.02	49.8	5	23	71	0
Maoming (MM)	6.41	47.13	12	45	73	0
Jiangmen (JM)	4.63	45.6	9	44	80	7
Zhongshan (ZS)	3.38	44.94	3	10	85	5
Zhanjiang (ZJ)	7.36	44.42	8	45	71	0
Shantou (ST)	5.66	39.04	9	75	71	5
Zhaoqing (ZQ)	4.19	32.59	8	10	86	10
Jieyang (JY)	6.11	30.46	1	33	71	5
Qingyuan (QY)	3.89	24.61	2	17	78	0
Shaoguan (SG)	3.03	19.11	6	48	78	0
Yangjiang (YJ)	2.57	18.73	1	19	71	15
Meizhou (MZ)	4.38	17.2	9	18	72	0
Chaozhou (CZ)	2.66	15.67	2	15	71	0
Shanwei (SW)	3.02	15.66	1	4	184	5
Heyuan (HY)	3.11	15.65	0	19	72	7
Yunfu (YF)	2.55	13.36	7	5	89	0

Data

This graph presents the relationship between average daily mobile users and residents of cities in Guangdong in 2019. The size of each marker denotes GDP per capita in 2019.



Data

Random 1% sample of the total users of the largest telecom service provider in Guandong who satisfy the following conditions

- Adult users aged 21 to 60
- Lived in Guangdong province in 2018.
- Excluding college students (younger than 25 and live on a university campus).
- Result: 392,087 users.

Data

Information for those in the sample

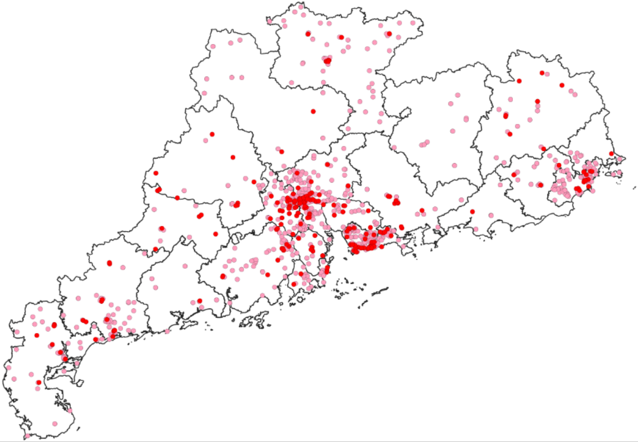
- Location records at five-minute intervals (in principle)
- Call records (encrypted IDs of the calling party and the receiving party, cell tower coordinates, date of calls, and call duration in seconds).
- App usage data (the time spent on each App in a month).
- Demographic information, such as age, gender, and the place where the phone number is registered.
- Restrictions on data usage: confidentiality agreement; access only in a company-owned data lab

Admission to top quality middle schools

- Primary school graduates have to take a city-level exam
- Those with the best scores (about 20% of the total) can enter an officially designated top middle school
- Other students enter a less selective school mostly based on proximity to homes
- Top middle schools ("provincial key schools") feature better teachers, faster pace of curriculum, better facilitates, academically stronger peers, etc.
- Admission to top quality middle schools is based on test sores. Rich families cannot buy their way in. No legacy admissions.

Top Quality Middle Schools

225 (out of 900+) middle schools are designated as "provincial-level key middle schools" based on a combination of absolute quality and relative quality.



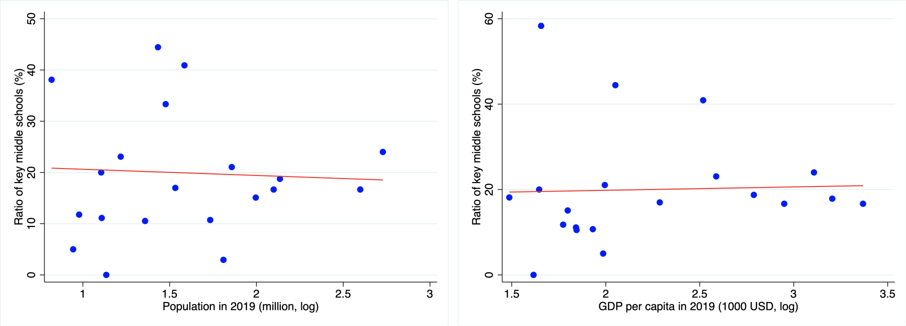
Key Middle Schools

The log number of key middle schools rises linearly with the log local population.

	(1)	(2)
	No. of key schools	
Population in 2019 (million, in log)	0.86**	-0.08
	(0.30)	(2.32)
Population squared term		0.30
		(0.73)
GDP per capita in 2019 (1,000 USD, in log)	0.44	1.27
	(0.28)	(3.05)
GDP per capita squared term		-0.20
		(0.67)
Observations	21	21
R-squared	0.46	0.40

Key Middle Schools

The ratio of top quality to all middle schools is uncorrelated with with log local population or log GDP per capita.



Identifying strategy

Identifying families with a child going from a primary school before the summer to a middle school in the fall (2019, 20, 21):

- Background: cell phone sub-stations are numerous, with usually a unique sub-station for a school
- No calls to people located in any middle school from January to June;
- Multiple calls (1 or more calls/month) to a middle school area from September to December (outside lockdown period);
- 1 or more calls/month to a primary school area from January to June outside lockdown);
- No calls to the same primary school area from September to December.

Our Sample

We identify 12,406 families with a child who moves from a primary school in the spring to a middle school in the fall. 20% (or 2,605) entered a top-quality middle schools.

Table: Family characteristics of our analysis sample

Variable	Mean	Std. Dev.	N
Average parent age (years)	39.16	12.40	12,406
Housing prices (million RMB)	2.69	5.79	12,406
Monthly high-end mall visits	2.12	1.96	12,406
Car owner (=1)	0.07	0.26	12,406
Average number of phones of parents	1.06	0.24	12,406
Average phone price category of parents	2.72	1.39	12,406

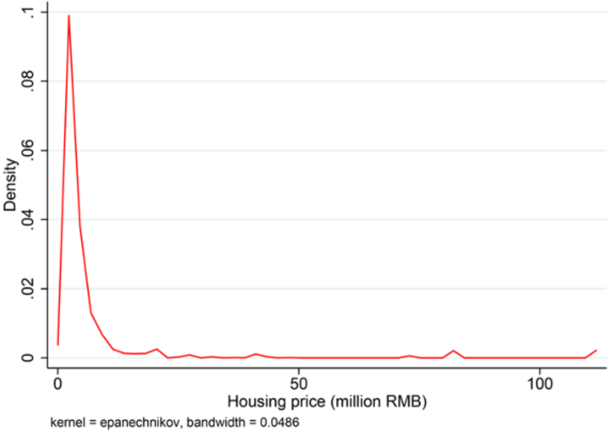
Proxy for family Wealth

= **average housing value in 2018** in the residential neighborhood (or cell tower area), usually a narrow area or a group of adjacent buildings

- Determine residential location: spending 5 or more hours during 10 pm - 7 am for 15 or more days/month for 10 or more months in 2018.
- Such geolocation information recorded for all mobile networks (2G, 3G, or 4G). The average coverage radius of 4G networks is 0.2-1.2 km.
- Using coordinates of each neighborhood to match the average home values in 2018 from Soufun.com (house transaction database).
- Maintained assumption: rank of family wealth = rank of average local home value

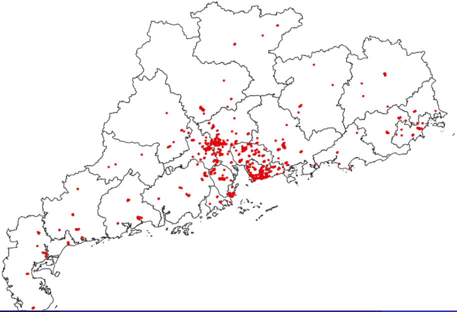
Caveat: cannot distinguish between homeowners and renters.

Distribution of home price per squared meters

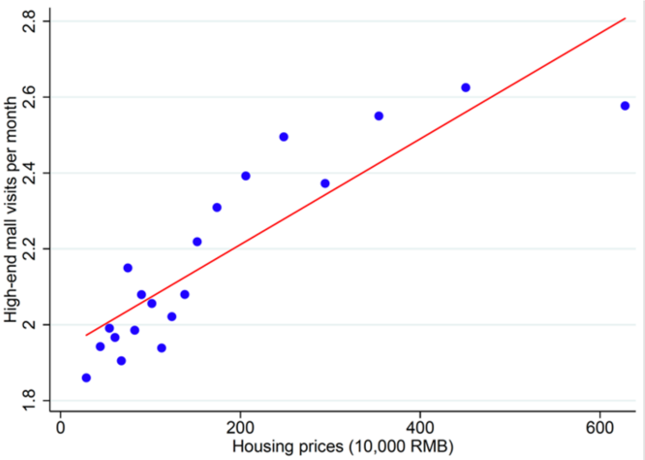


Validating the Proxy for Family Wealth by digital footprints

- Frequency of visits to high-end shopping centers -those with dedicated stores for expensive apparel/footwear brands or jewellery stores
- Louis Vuitton, Cartier, Hermes, Gucci, Rolex, Nike, Adidas, HM, Zara, and Uniqlo
- Jewellery: Chow Tai Fook, Chow Sang Sang, Lao Feng Xiang, Lukfook, TSL, CHJ, Chow Tai Seng, Tesiro, I DO, and Kneerworld.
- 1,128 high-end shopping centers across the sample cities.
- A visit = a trip there between 1-4 hours



Validating with digital footprints



Correlations between proxies for wealth

	Housing prices	High-end mall visits	Car owner	Number of Phones	Phone price category
Housing values	-				
	-				
High-end mall visits	0.051***	-			
	(0.000)	-			
Car ownership	0.001	0.160***	-		
	(0.544)	(0.000)	-		
Number of phones	0.003*	0.002	0.032***	-	
	(0.070)	(0.302)	(0.000)	-	
Phone price category	0.006***	0.002	0.096***	0.153***	-
	(0.002)	(0.202)	(0.000)	(0.000)	-

Our Sample

We identify 12,406 families with a child who moves from a primary school in the spring to a middle school in the fall. 20% (or 2,605) entered a top-quality middle schools.

Table: Family characteristics of our analysis sample

Variable	Mean	Std. Dev.	N
Average parent age (years)	39.16	12.40	12,406
Housing prices (million RMB)	2.69	5.79	12,406
Monthly high-end mall visits	2.12	1.96	12,406
Car owner (=1)	0.07	0.26	12,406
Average number of phones of parents	1.06	0.24	12,406
Average phone price category of parents	2.72	1.39	12,406

School Closure and Education Gap

Specification: family wealth, school closure, and top school admission

$$E_{ict} = \alpha_1 W_{i,2018} + \alpha_2 W_{i,2018} \times SC_{ict} + \alpha_3 SC_{ict} + \delta_c + \eta_t + \epsilon_{ict} \tag{2}$$

E_{ict} = 1 if student i is admitted to a top middle school in city c and year t.

$W_{i,2018}$ = family wealth in 2018 for student i.

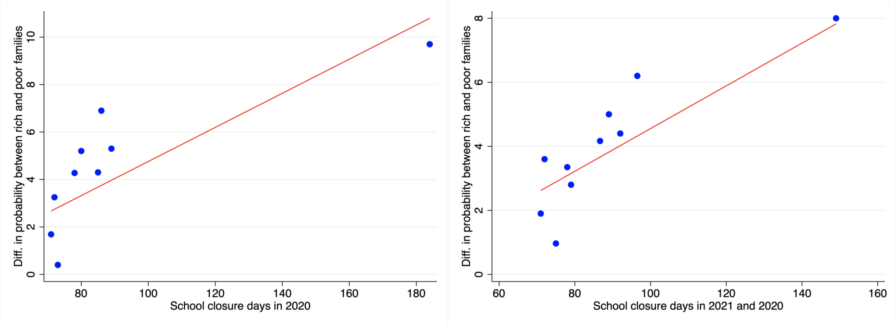
SC_{ict} = number of days school closures before summer in city c and year t.

δ_c : city fixed-effect.

η_t : year fixed-effect.

Standard errors clustered by city: wild-boostrapping method (Wu, 1986; Davison and Flachaire, 2008)

Binscatter polts



(a) cohort 2020 vs 2019

(b) cohort 2021 vs 2019

The y-axis in panel A (B) denotes the difference in difference in probability between rich and poor families between 2020 (2021) and 2019 cohorts.

Relative Wealth Within a City

	Entering key middle schools (=1)				
	(1)	(2)	(3)	(4)	(5)
Home (1 million RMBs)	0.006*		0.009***	0.007***	
	(0.003)		(0.001)	(0.001)	
Home (0-25% within city)		-0.036*			-0.030*
		(0.020)			(0.017)
Home (50-75% within city)		0.028*			0.045**
		(0.016)			(0.021)
Home (75-100% within city)		0.092**			0.091***
		(0.045)			(0.029)
School-closure days (10 days)			0.006	-0.009***	
			(0.012)	(0.002)	
School closure days (10 days)				0.005***	
* Home (1 million RMB)				(0.001)	
School closure days (10 days)					-0.002*
* Home (0-25% within city)					(0.001)
School closure days (10 days)					0.002*
* Home (50-75% within city)					(0.001)
School closure days (10 days)					0.004**
* Home (75-100% within city)					(0.002)
Sample	2019	2019	2019-2021	2019-2021	2019-2021
Baseline mean	0.217	0.217	0.217	0.217	0.217
City and year FEs	Yes	Yes	Yes	Yes	Yes
Observations	3,925	3,925	12,406	12,406	12,406

Effect size

	(1)	(2)	(3)
Probability of entering key middle school	Bottom quartile	Top quartile	Diff (2)-(1)
Unconditional mean (2019 sample)	17.7%	30.5%	12.8%
For cities experiencing 70 school-closure days	16.3%	33.3%	17.0%
For cities experiencing 180 school-closure days	14.1%	37.7%	23.6%

Relative Wealth Within a City

	Ranking of middle schools (within city percentile%)				
	(1)	(2)	(3)	(4)	(5)
Home (1 million RMBs)	0.011**		0.014***	0.010***	
	(0.005)		(0.002)	(0.002)	
Home (0-25% within city)		-0.056***			-0.049***
		(0.017)			(0.018)
Home (50-75% within city)		0.037*			0.033
		(0.020)			(0.022)
Home (75-100% within city)		0.105***			0.098***
		(0.032)			(0.034)
School-closure days (10 days)			-0.001	-0.015***	
			(0.007)	(0.005)	
School closure days (10 days)				0.008**	
* Home (1 million RMB)				(0.004)	
School closure days (10 days)					-0.006**
* Home (0-25% within city)					(0.003)
School closure days (10 days)					0.004*
* Home (50-75% within city)					(0.002)
School closure days (10 days)					0.008***
* Home (75-100% within city)					(0.003)
Sample	2019	2019	2019-2021	2019-2021	2019-2021
Baseline mean	0.513	0.513	0.513	0.513	0.513
City and year FEs	Yes	Yes	Yes	Yes	Yes
Observations	3,925	3,925	12,406	12,406	12,406

2020 vs. 2021

	Entering key middle schools (=1)			
	(1)	(2)	(3)	(4)
Home (1 million RMBs)	0.005**		0.008**	
	(0.002)		(0.004)	
Home (0-25% within city)		-0.026*		-0.032*
		(0.015)		(0.019)
Home (50-75% within city)		0.032		0.054*
		(0.020)		(0.030)
Home (75-100% within city)		0.085**		0.100**
		(0.035)		(0.041)
School-closure days (10 days)	-0.007**		-0.010***	
	(0.003)		(0.003)	
School closure days (10 days)	0.004**		0.007**	
* Home (1 million RMB)	(0.002)		(0.003)	
School closure days (10 days)		-0.002*		-0.003**
* Home (0-25% within city)		(0.001)		(0.001)
School closure days (10 days)		0.001*		0.003*
* Home (50-75% within city)		(0.000)		(0.002)
School closure days (10 days)		-0.003*		-0.005**
* Home (75-100% within city)		(0.002)		(0.002)
Sample	2019&2020	2019&2020	2019&2021	2019&2021
Baseline mean	0.217	0.217	0.217	0.217
City and year FEs	Yes	Yes	Yes	Yes
Observations	8,279	8,279	8,052	8,052

Parent Digital App Usage

Are there different patterns of app usage by the parents? Do they matter?
Parent app usage when their child in last semester of primary school

$$P_{ict} = \alpha_1 W_{i,2018} + \alpha_2 W_{i,2018} \times SC_{ict} + \alpha_3 SC_{ict} + \delta_c + \eta_t + \epsilon_{ict} \quad (3)$$

P_{ict} : hours spent on mobile Apps or average daily home time during weekdays in month t.

SC_{ict} : the number of days of Covid-induced school closures in each month t.

θ_c : individual fixed-effect.

λ_t : year-by-month fixed-effect.

Standard errors clustered by city: wild-boostrapping method (Wu, 1986; Davison and Flachaire, 2008)

Table: Hours allocated per month by parent characteristics

	Panel A	Panel B: By gender		Panel C: By housing prices		
	Full sample	**Mothers**	**Fathers**	**Below 1**	**1-3**	**Above 3**
Observations	568,571	314,441	254,130	260,895	224,650	83,026
No. of users	17,768	8,471	9,297	8,692	7,019	2,057
Mobile time	93.7	88.4	96.5	87.5	93.3	104.9
Child-education	3.1	4.2	2.2	2.5	3.0	4.0
	(70.4)	(85.8)	(55.8)	(66.0)	(67.8)	(82.8)
Game	14.7	13.2	15.8	14.1	14.7	15.7
	(53.5)	(50.5)	(55.7)	(54.6)	(53.5)	(51.7)
Video	6.4	6.3	6.4	5.8	6.3	7.4
	(21.9)	(22.0)	(21.8)	(20.9)	(21.9)	(23.8)
Social media	59.1	54.0	63.1	55.6	59.2	66.0
	(61.5)	(59.2)	(63.2)	(59.6)	(61.7)	(64.4)
Others	10.4	10.7	9.0	9.5	10.1	11.8
	(10.6)	(7.6)	(12.4)	(10.4)	(10.4)	(11.0)

Digital App Usage

Table: Effect of Covid-induced school closure days on App usage

	(1) Child edu	(2) Games	(3) Video	(4) Social media	(5) Other	(6) Daily home time (hou
Panel A: Mother sample						
No. of school closure days (10 days)	0.15**	0.04	0.10	0.04	0.07	1.07*
	(0.07)	(0.05)	(0.11)	(0.05)	(0.06)	(0.56)
No. of school closure days (10 days)	0.39***	-0.12	0.03	-0.12	0.06	0.25**
* Housing prices (1 million RMB)	(0.12)	(0.10)	(0.07)	(0.10)	(0.06)	(0.11)
Baseline mean	3.8	12.7	6.1	54.9	10.6	12.3
Individual and year-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138,752	138,752	138,752	138,752	138,752	138,752
Panel B: Father sample						
No. of school closure days (10 days)	0.04	0.26***	0.19	0.26***	0.19**	0.53
	(0.07)	(0.09)	(0.14)	(0.09)	(0.09)	(1.46)
No. of school closure days (10 days)	0.06*	-0.12**	0.09	-0.12**	0.05	-0.32
* Housing prices (1 million RMB)	(0.04)	(0.06)	(0.07)	(0.06)	(0.05)	(0.71)
Baseline mean	2.2	15.8	6.2	64.5	9.2	11.2
Individual and year-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	135,219	135,219	135,219	135,219	135,219	135,219

Conclusion

- While rich children are more likely to go to a top middle school even without school closure, we find that their advantage increases with the duration of school closure.
- Different app usage habits (e.g., the relative time spent on education apps versus games) during school closure might be a contributing factor to the widening educational gap.

Thank You!