October 2022

What Do Changes in State Test Scores Imply for Later Life Outcomes?

Thomas J. Kane
Elena Doty
Tyler Patterson
Douglas O. Staiger

DARTMOUTH

Center for Education Policy Research
HARVARD UNIVERSITY

SUGGESTED CITATION
Kane, T., Doty E., Patterson, T., Staiger, D., (2022) What Do Changes in State Test Scores Imply for Later Life Outcomes? Cambridge, MA: Center for Education Policy Research, Harvard University

This research was supported by a grant from The Walton Family Foundation.

Abstract

In the three decades before the pandemic, mean achievement of U.S. 8th graders in math rose by more than half a standard deviation on the National Assessment of Educational Progress (NAEP). Between 2019 and 2022, U.S. students had forfeited 40 percent of that rise. To anticipate the consequences of the recent decline, we investigate the past relationship between NAEP scores and students' later life outcomes by year and state of birth. We find that a standard deviation improvement in a birth cohort's 8th grade math achievement was associated with an 8 percent rise in income, as well as improved educational attainment and declines in teen motherhood, incarceration and arrest rates. If allowed to become permanent, our findings imply that the recent losses would represent a 1.6 percent decline in present value of lifetime earnings for the average K-12 student (or $19,400), totaling $900 billion for the 48 million students enrolled in public schools during the 2020-21 school year.

INTRODUCTION

In the three decades before the pandemic, mean math achievement of U.S. 8th graders on the National Assessment of Educational Progress (NAEP) had improved substantially (by slightly more than half a standard deviation.)[1] Even before the recent losses, the biennial release of state rankings on the NAEP had become a major news event, garnering headlines around the country. But do the scores provide a valid signal of students' future social and economic prospects—or do they simply reflect test-taking skills?

It is a timely question, given that mean 8th grade math achievement declined by .2 standard deviations between 2019 and 2022, effectively forfeiting forty percent of the prior increase. Although the long-term consequences that decline remain to be seen, we investigate the changes in later life outcomes by year and state of birth which accompanied past changes in achievement on the NAEP tests.

To do so, we exploit state-level variation in achievement gains since 1990, when the NAEP first began measuring state differences in achievement. In a subset of states, NAEP scores rose sharply between 1990 and 2019. For instance, the average student in North Carolina improved by .94 standard deviations in 8th grade math achievement—an improvement roughly equal to the national black-white gap in 1990 (.91 standard deviations). Meanwhile, the gains in some other states, such as Iowa, were much smaller (.10 standard deviations). We use these state-level differences in achievement gains on the NAEP along with outcomes by year and state of birth in the American Community Survey to estimate the association between past achievement increases and later-life outcomes. We then use those relationships to anticipate the future implications of the recent declines in achievement for current K-12 cohorts.

We find that a standard deviation rise in 8th grade math achievement is associated with an 8 percent rise in adult's earned income, as well as improvements in educational attainment and declines in unemployment, teen motherhood, incarcerations and arrests. The estimates are robust to include controls for age, parental education and race/ethnicity. We adjust for unmeasured social background factors with fixed effects for state of birth and for region of birth by year of birth. We also adjust for changing labor market conditions by including fixed effects for state of residence by year. By our estimates, the recent decline would represent a 1.6 percent decline in present value of lifetime earnings, as well as a decline in high school graduation and increase in teen motherhood, incarceration rates and arrest rates—if it were to become permanent.
When spread across the 48 million public school students in the U.S. the estimated loss would total $1 trillion.

RELEVANT LITERATURE

There is a body of large literature measuring the relationship between adult earnings and pre-labor market achievement test scores for individuals. For instance, Neal and Johnson (1996) find that a standard deviation in AFQT

[1] The federal government invests more than $175m annually on the assessments. National Academies of Sciences, Engineering and Medicine (2022).

scores was associated with roughly a 20 percent difference in earnings at age 26-29 for both men and women. Murnane et al. (2000) find that a standard deviation in 10th grade math scores was associated with a 12 percentage point difference in earnings at age 31. More generally, early life test scores have been linked to a wide range of later life outcomes for individuals, including educational attainment, teenage childbearing and illegal behaviors (Goldhaber and Ozek, 2019).

Achievement test scores are likely correlated with other factors influencing earnings, such as family background. However, policies such as school spending and school accountability clearly play a role. In a study of the impact of school accountability pressure on the subsequent achievement and earnings of students in Texas, Deming et al. (2016) found that the policy change boosted math achievement for low-achieving students by .2 standard deviations and raised earnings at age 25 by 2.1 percent. If the impact of school accountability operated primarily through boosting academic achievement, the implied instrumental variable estimate would be similar to the cross-sectional estimates, about 11 percent (2.1/.2). The Deming et al. study is especially relevant for our current study, given that state accountability policies have been shown to be an important driver of the improvement in mean scores on the NAEP during the 1990's (Dee and Jacob, (2011), Hanushek and Raymond, (2005)).

There is a related literature on the relationship between international differences in economic growth and educational attainment/ cognitive test scores (Barro, (1991), Hanushek and Woessman, (2008)). One obvious challenge is the potential endogeneity of growth with educational attainment and achievement (Bils and Klenow, 2000). We face a similar challenge in studying the relationship between achievement growth in a state and a birth cohort's income and educational attainment. Thus, we use a number of statistical controls in attempt to address endogeneity: relying on changes in achievement within state of birth, controlling for labor market conditions by including fixed effects by state of residence by year, controlling for race and parental education, and region by year of birth.

DATA

To measure state differences in student achievement over time, we rely on the Main NAEP.[2] The assessment collects data from a representative sample of more than 100 public schools in each participating state, with a total of roughly 4,800 schools and 125,000 students annually.[3] Prior to 2002, states could opt out of participating, and the number of participating states ranged from 38 to 46, depending upon the subject and grade. Since 2003, all states have participated and the NAEP has been administered every other year, in both reading and mathematics.

For this analysis, we focus on the 8th grade mathematics assessment, which was administered in 1990, 1992, 1996, 2000 and biennially from 2003 to 2019. Due to the pandemic, the 2021 assessments were postponed and the tests were administered January-March 2022. We focus on math achievement because prior research has found it to be is more correlated with earnings than reading or vocabulary (Murnane, Willett and Levy,

[2] In contrast to the Long-term Trend NAEP, which tracks national trends, the Main NAEP was designed to yield state-level estimates.
[3] Public school enrollment represented 90.2 percent of K-12 enrollment in the U.S. in 2019. https://nces.ed.gov/programs/digest/d21/tables/dt21_205.10.asp?current=yes

1995). Moreover, we use 8th grade math achievement, since 4th grade scores were not available by state for 1990 and because 8th grade scores allow us to include older birth cohorts.

Specifically, we use the mean score of 8th graders in a state as an estimate of the mean achievement of those born in the state 13 years before.[4] We impute scores for missing years using the average of scores from years t-1 and t+1 if they are available. (If there is more than a one year gap, we treat achievement as missing for the intervening years.) We standardize scores in all years using the national mean and standard deviation from 1992.

In Figure 1, the height of the bars portrays the change in mean 8th grade math scores by state between 1990 and 2019. Every state's mean score in 2019 exceeded its 1990 mean score, although the magnitude of the improvement varied—from .10 s.d. in Montana to .92 s.d. in North Carolina.[5] The black arrows portray the change in scores between 1992 and 2022 – so that the difference between the height of the bar and the corresponding black arrow is the decline in 8th grade math scores that occurred during the pandemic. For example, the 2019-22 losses were nearly as large as the improvement over the prior three decades in Oklahoma, New Mexico and Missouri. Students in five states (Oregon, North Dakota, Maine, Iowa and Montana) actually scored below their 1990 counter-parts.

In addition to using mean achievement by state, we use the student-level microdata from the NAEP to adjust state achievement measures for changes in race/ethnicity and parental education of tested students. To do so, we estimate the following model:

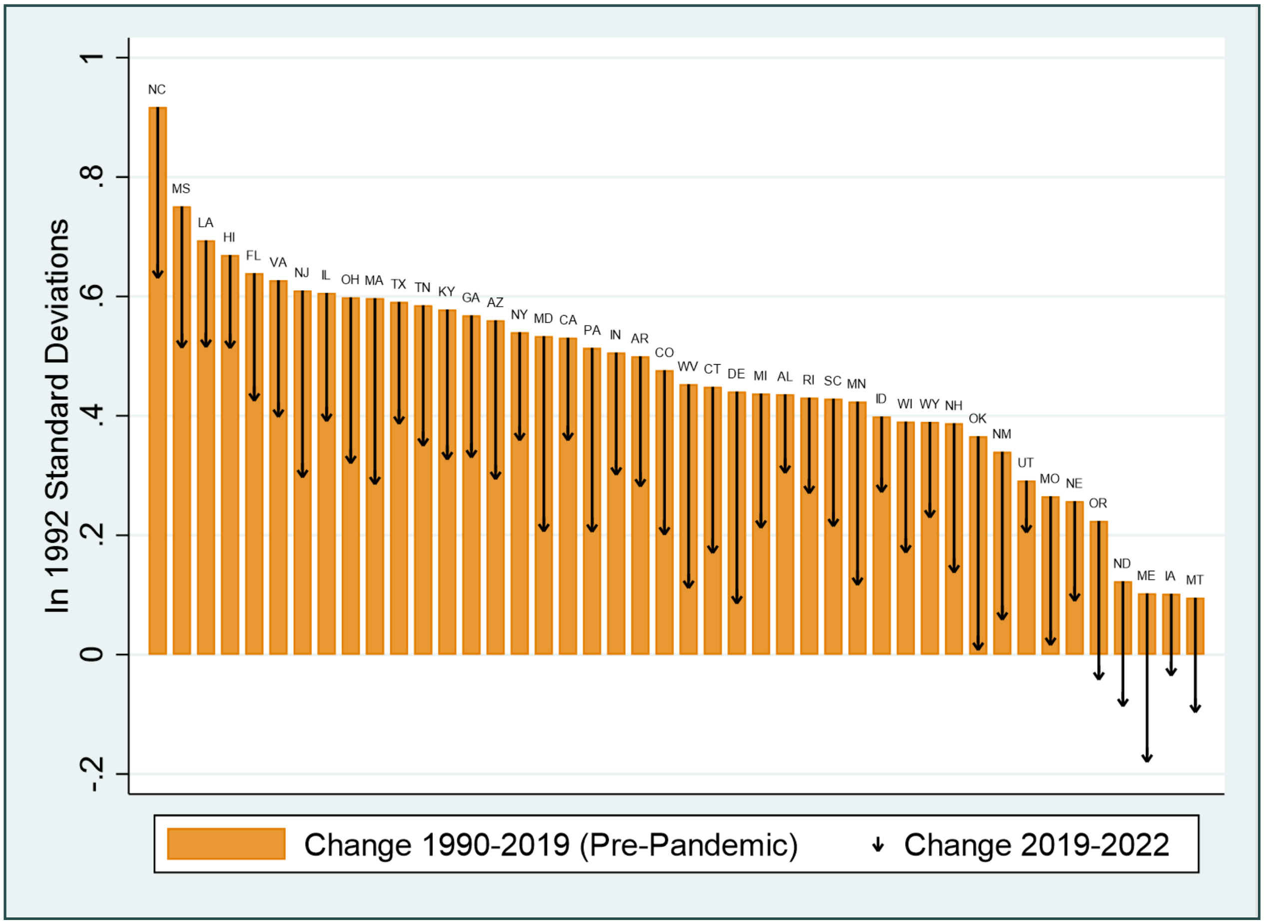
$$(1)\ S_{ijc} = \beta_0 + \boldsymbol{Race_i}\beta_1 + \boldsymbol{ParentEd_i}\beta_2 + \delta_{jc} + \varepsilon_{ijc}$$

where $Race_i$ is a vector of indicators of student race/ethnicity[6] and $ParentEd_i$ is a vector of indicators of student-reported parental education[7]. We refer to the state by cohort effects from equation (1), $\hat{\delta}_{jc}$, as "adjusted math achievement."

To measure later life outcomes, we use the 1 percent sample from the 2000 Decennial Census and the American Community Survey (ACS) from 2001 through 2019 (Ruggles et al., 2022).

[4] We believe this to be a reasonable assumption given that the typical 8th-grader enters at age 13 and 80 percent of 13 year olds were still living in their state of birth in 2000.
[5] In Iowa, the mean score declined by .05 s.d. from 1992 to 2019, though it improved by .15 s.d. from 1990 to 1992.
[6] White, Black, Hispanic, Asian/Pacific Islander, Native American/Alaskan Native and Other/unclassified.
[7] We included a set of "dummies," 10 indicators for mother's and father's education. For each parent, students were asked to choose from five categories: did not finish high school, high school graduate, some education after high school, graduated from college and "I don't know."

Figure 1. Changes in 8th grade Math Achievement Before and After Pandemic by State



Note: Based on mean 8th grade math achievement in National Assessment. For the seven states that had a valid score for 1992, but not 1990 (Massachusetts, Maine, Missouri, Mississippi, South Carolina, Tennessee, and Utah), we report the change since 1992. Six states had neither a 1990 nor a 1992 NAEP score (Alaska, Kansas, Nevada, South Dakota, Vermont, and Washington) and are excluded from the graph. Mean achievement had increased in all states over the 29 years preceding the pandemic. Every state lost ground between 2019 and 2022, with five states scoring below their 1990 (or 1992) means. On average, the U.S. forfeited 41 percent of the pre-pandemic improvement (1990 to 2019) between 2019 and 2022.

In addition to providing data on earned income, occupation, educational attainment, college enrollment, teen motherhood, Food Stamp (SNAP) receipt, home ownership and incarceration, the Census and ACS provide information on year and state of birth.[8]

We use state of birth as a proxy for the state where the student was residing in grade 8. In the 2000 Census, 80 percent of U.S. born 13 year-olds were living in their state of birth. Because the District of Columbia was an outlier (21 percent of those born in DC were residing in the district at age 13), we drop those born in D.C. Prior work using the ACS to link later life outcomes to state policies has found that using state of birth introduces little bias (Lovenheim & Willen, 2019).

In addition to the Census and American Community Survey, we use estimates of violent and property crime arrests by state of occurrence, year and age from the the FBI Uniform Crime Report Arrest Master File (materials from Donohue and Levitt, 2020).[9]

CHANGES IN ACHIEVEMENT AND LATER LIFE OUTCOMES BY STATE OF BIRTH

In Figure 2, we report scatterplots of changes in outcomes against changes in math achievement by state of birth. For each outcome, we use the maximum span of birth cohorts for which we are able to observe the outcome. For instance, in the top left panel, we report the change in mean log earned income (which we measure for those aged 28 and above) against the change in achievement for those born in 1977 (age 13 in 1990) and those born in 1990 (age 13 in 2003).[10] The bivariate slope coefficient on the long difference is 0.079, similar to estimates of the impact of test scores on earnings based on individual-level data summarized above.

In the remaining panels of Figure 2, we report similar graphs for five additional outcomes: the change in high school graduation rates (for those aged 23 and above), the change in college enrollment (for those aged 18 to 24), the change in teen motherhood (for females aged 18 to 30), the change in rates of institutionalization (for males between the ages of 18 and 24), and the change in arrest rates for violent crimes (for males aged 18 to 24). The states with the largest improvements in achievement also saw larger improvements in high school graduation and college enrollment, as well as bigger decreases in teen motherhood, rates of institutionalization and arrests for violent crimes.

INCOME SPECIFICATIONS

We investigate the robustness of these bivariate relationships by introducing a variety of statistical controls. We start by estimating the following linear model of log of earned income Ln Y_{ijct}:

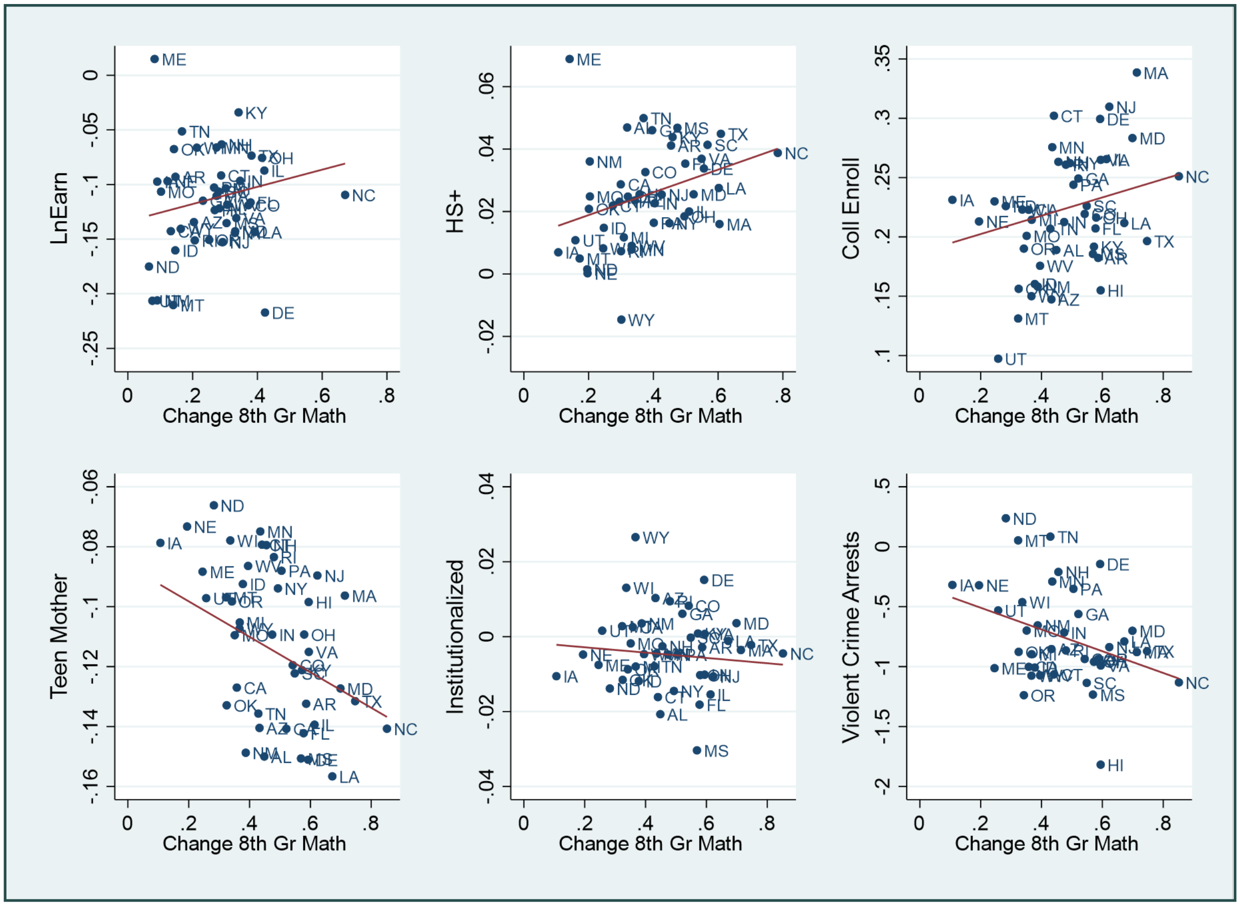
$$(2) \quad Ln\, Y_{ijct} = \gamma_o + \gamma_1 Score_{jc} + \gamma_2 Gender_i + \boldsymbol{Race_i}\gamma_3 + \theta_j + \mu_c + \pi_{age} + \tau_t + \varepsilon_{ijt}$$

[8] In the 2021 ACS, 98 percent of the 18-24 year old population that were in institutional group quarters were in adult correctional facilities or juvenile detention. As a result, we use an indicator of residence in institutional group quarters as a proxy for incarceration. (US Bureau of the Census, Table S2603)

[9] https://works.bepress.com/john_donohue/192/

[10] Note that for this graph the 1977 cohort and 1990 cohort are observed at different ages in the ACS, but this is the same for every state so that the long differences still capture how changes in cohort earnings are associated with changes in cohort scores. In the regressions we control for age at which the cohort was observed.

Figure 2. Within-State Changes in Birth Cohort Outcomes and Corresponding Changes in 8th Grade Math Achievement



Note: We use the longest span in birth year possible for each of the outcomes. For instance, we measure log earnings for those aged 28 and above. Thus, the earliest birth cohort, the 8th graders in 1990, was born in 1977 and was age 42 in 2019. The latest birth cohort for whom we could measure earnings was born in 1990 and were in 8th grade in 2003. The cohorts used for high school completion (which we measure at ages 23 to 42) were born between 1977 and 1994; college enrollment (which we measure between ages 19 and 24) was measured for the 1977 to 1998 birth cohorts; teen motherhood (which we measure between ages of 18 to 30) includes the 1977 to 1998 birth cohorts; institutionalization (which we measure between ages 18 and 24) includes the 1977 to 1998 birth cohorts; violent crime arrests (which we measure for those aged 15 to 24) includes the 1977 to 1998 birth cohorts.

Where i subscripts the individual, j the state of birth, c the year of birth and t the year of the outcome is measured. $\overline{Score}_{jc}$ is the mean score on the 8th grade math assessment. Equation (2) includes controls for gender and race, as well as fixed effects for state of birth, birth cohort, year and year of age. Standard errors are calculated while clustering at the state of birth by cohort level. In column (1), the coefficient on mean 8th grade math score is .111.

Although we are conditioning on each sample member's race/ethnicity in equation (2), neither the ACS nor the Census provides a measure of family background (such as parental education or family income) for adults. Yet, if family resources were rising and had a direct positive effect on both achievement and income, the coefficient on mean achievement could be biased upward. Thus, in column (2), we replace the mean score, $\overline{Score}_{jc}$, with the state-year means adjusted by race and parental education (i.e. $\hat{\delta}_{jc}$ from equation (1)). The coefficient on adjusted math achievement is somewhat smaller, .096.

While test scores have been adjusted for parental education, a student's parental education may have had a direct effect on later life outcomes. Moreover, a larger concentration of college graduates among parents may have broader effects on the cohort—e.g. leading to investments by parents in social networks or other opportunites (e.g. extra-curricular activities) not reflected in test scores. Thus, in column (3), we add a control for the percent of students in the birth cohort and state who had a resident parent with a college degree (associate's, bachelor's or higher).[11] The coefficient falls slightly to .074.

In column (4), we add controls for state of residence, to complement the controls for state of birth. The coefficient on adjusted math achievement remains statistically significant at .062.

To ensure that the results are not simply being driven by differential trends for students born in different regions of the country (for example, perhaps economic development was improving both achievement and incomes in the South), we add controls for nine geographic divisions interacted with single year of birth in column (5). (Column (5) will serve as our preferred specification for later outcomes.)[12] The coefficient on adjusted test scores remains significant at .083. (Recall that some southern states, such as NC, saw much larger improvements in achievement than others, such as Alabama.)

In column (6), we add "dummies" for state of residence interacted by year to account for changing labor markets. Again, the coefficient on adjusted test scores remains significant at .077. In column (7), we add separate time trends for each state of birth. Since the adjusted scores vary by state of birth and year, the only

[11] We estimate parental education in the state and cohort using the 1990 and 2000 decennial censuses and the 2009 ACS, which provided large samples of students in each birth cohort in each state. We normalize scores for each birth cohort, since older cohorts with older parents may report additional schooling.

[12] The nine census divisions include New England (CT,ME,MA,NH,RI,VT), Mid-Atlantic (NJ, NY, PA), East North Central (IN, IL, MI, OH, WI), West North Central (IA,KS,MN, MO, NE, ND, SD), South Atlantic (DE, DC, FL, GA, MD, NC, SC, VA, WV), East South Central (AL, KY, MS, TN), West South Central (AR, LA, OK, TX), Mountain (AZ, CO, ID, NM, MT, UT, NV, WY) and Pacific (AL, CA, HI, OR, WA).

remaining variation is in the differing timing of achievement increases. The point estimate remains similar, at .087, but the standard errors nearly double. Thus, much of the variation driving the precision of our estimates comes from differential state trends in NAEP scores (the long-difference variation seen in Figure 2) rather than year to year variation.

OTHER LIFE OUTCOMES FROM THE AMERICAN COMMUNITY SURVEY

In Table 2, we report the relationship for additional outcomes in the Census and American Community Survey. Rather than replicate all the columns in Table 1, we report results for our preferred specification (column 5, although the results are similarly robust to alternative specifications.) We find that those born in states with larger increases in 8th grade math achievement saw larger increases in high school graduation and an increase in college enrollment. With regard to labor force participation, birth cohorts in those states were less likely to be unemployed and worked longer hours. They were more likely to be either working full time or enrolled in college when observed between the ages of 18 and 24. Girls born in those states were less likely to become teen mothers; boys were less likely to be incarcerated at age 18 to 24.

The relationships for several other outcomes were not statistically significant (specifically, BA degree completion, home ownership, occupational prestige and marital status). BA degree completion was the outcome most related to parental education. And one coefficient, on Food Stamp (or SNAP) benefit receipt, had an anomalous (positive) sign.

ARRESTS

In Table 3, we report the relationship between the log of the number of arrests by single year of age, state and year and the measure of math achievement adjusted for parental education and race ($\hat{\delta}_{jc}$). As in the previous analyses, we include a measure of the percentage of parents in the birth cohort who were college graduates as well as fixed effects for states, age by year and geographic Census division by year of birth. Unlike the ACS and Census, arrests are only available by state of occurrence, not by state of birth. Thus, unlike in Tables 1 and 2, we assign math achievement by state of occurrence and limit the outcome to young adults, aged 15-24.[13] A one standard deviation increase in 8th grade math achievement is associated with a 36 percent decline in violent crime arrests and a somewhat smaller (20 percent) decline in property crime arrests.[14]

IMPLICATIONS OF ACHIEVEMENT LOSSES

The average math achievement of U.S. 8th graders nationally improved by half a standard deviation (18 points) between 1990 and 2007 (the year today's 28 year olds would have been in 8th grade). Our results imply that the annual incomes of U.S. 28 year olds are 4.2 percent higher as a result, while teen motherhood rates and male incarceration

[13] In the 2009 ACS, 73 percent of U.S.-born individuals between the ages of 18 and 24 were still living in their state of birth.
[14] If β is the coefficient on test scores and the dependent variable is the natural log of arrests, then the implied percentage change in arrests per unit change in test scores is $e^{\beta}—1$.

Table 1. The relationship between log income and math achievement by state birth cohort

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Math Achievement	0.111***						
	(0.022)						
Adjusted Math Achievement		0.096***	0.074**	0.062*	0.083**	0.077*	0.087
		(0.025)	(0.024)	(0.024)	(0.030)	(0.033)	(0.058)
Either Parent College Graduate			0.175***	0.180***	0.105	0.112*	0.038
			(0.037)	(0.037)	(0.057)	(0.056)	(0.061)
Fixed Effects:							
Gender	X	X	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X
Year	X	X	X	X	X	X	X
Year of Birth	X	X	X	X	X	X	X
State of Birth	X	X	X	X	X	X	X
State of Residence				X	X	X	X
Division of Birth X YOB					X	X	X
State of Resid. X Year						X	
Trends by YOB and State of Birth							X
N	906,767	906,608	906,608	906,608	906,608	906,608	906,608

Note: Dependent variable is natural log of income for those aged 28 and above. The sample is drawn from the 2000 Decennial Census (1 percent sample) and the 2001-2019 American Community Survey from IPUMS. We measure math achievement for each birth cohort in a state using the mean achievement of 8th graders on the National Assessment of Educational Progress 13 years later. “Adjusted math achievement” has been adjusted for race/ethnicity and parental education. (See equation (1).) Column (5) reflects our preferred specification. Standard errors are calculated after clustering by year and state of birth.

Table 2. The relationship between other later life outcomes and math achievement by state birth cohort

	Adjusted Math Score	% Either Parent College Graduate	Age when Observed	Birth Cohorts Used	Sample Size
Ln(Earned Income)	0.083**	0.101	28-42	1977-1992	906,608
	(0.030)	(0.058)			
Occupational Prestige Score	0.126	-0.786	28-42	1977-1992	1,177,134
	(0.401)	(0.742)			
Educational Attainment:					
HS+	0.014*	0.011	23-42	1977-1996	2,270,247
	(0.006)	(0.013)			
Some College +	0.018	0.064**	23-42	1977-1996	2,270,247
	(0.010)	(0.025)			
BA +	0.011	0.112***	23-42	1977-1996	2,270,247
	(0.010)	(0.024)			
Graduate Degree	0.006	0.031**	23-42	1977-1996	2,270,247
	(0.005)	(0.011)			
Unemployed	-0.020*	0.007	28-42	1977-1992	1,031,871
	(0.008)	(0.016)			
Receiving Food Stamps/SNAP	0.027*	-0.009	28-42	1977-1992	1,241,219
	(0.014)	(0.027)			
Usual Hours Worked	0.737*	-0.831	28-42	1977-1992	1,002,229
	(0.316)	(0.698)			
Ever Married	-0.022	0.014	28-42	1977-1992	1,237,417
	(0.016)	(0.031)			
Own Home	-0.011	0.000	28-42	1977-1992	1,193,585
	(0.017)	(0.033)			
Teenage Mother	-0.018*	0.035*	18-30	1977-2002	1,583,294
	(0.009)	(0.014)			
Institutionalized/Incarcerated	-0.017***	0.010	18-24	1977-2002	1,076,101
	(0.004)	(0.006)			
Enrolled in College	0.051***	0.045	19-24	1977-2000	1,770,373
	(0.012)	(0.024)			
Working FT/Enrolled in College	0.027**	0.015	19-24	1977-2000	1,677,135
	(0.009)	(0.018)			

Note: All of the above specifications also included fixed effects for age, state of birth, state of residence, census division of birth by year of birth and year. The sample is drawn from the 2000 Decennial Census (1 percent sample) and the 2001-2019 American Community Survey, downloaded from IPUMS. Standard errors are calculated after clustering by state and year of birth.

Table 3. Age Specific Arrests and 8th Grade Math Achievement

	Violent Crime	Property Crime
Adjusted Math Achievement	-0.357***	-0.201*
	(0.088)	(0.087)
Either Parent College Graduate	0.098	-0.116
	(0.178)	(0.163)
State	X	X
Age X Year	X	X
Geographic Division X Year of Birth	X	X
N	4,130	4,130

Note: The dependent variable is the log of arrests by state, age, and year. The data on arrests were provided by Donohue and Levitt (2021). "Adjusted Math Achievement" consists of state by year fixed effects, after using student-level data from the Main NAEP to adjust for parental education and student race-ethnicity. (See model (1).) Standard errors are calculated clustering by state and year of birth.

rates both declined by 0.9 percentage points. The improvements in income in North Carolina would have been even larger, over 7 percent, given the magnitude of their increases.

However, between 2019 and 2022 national mean 8th grade math achievement declined by .2 standard deviations, meaning U.S. students have forfeited forty percent of the improvement since 1990.[15] The decline in 4th grade math was similar, .16 standard deviations. Thus, if such losses were to become permanent, our results imply that a .2 standard deviation decline in achievement would be associated with a 1.6 percent decline in earnings for both cohorts of students.

A 1.6 percent decline in earnings may seem like a modest impact per student, but it adds up. Suppose we were to assume that the average public school student suffered a .2 standard deviation loss in achievement.[16] Following Krueger (2003), we use the 2019 American Community Survey to estimate the present value of lifetime earnings between ages 28 and 64. Valuing future earnings as of their current grade in 2020-21, we estimate an average present value of lifetime earning across grades of $1.2 million.[17] Thus, a 1.6 percent decline in future earnings would represent an $19,400 loss in present value for the average student who was in grades K-12 during the 2020-21 school year. When multiplied by 48 million (the number of students enrolled in public K-12 schools during 2020-21), the total loss becomes $900 billion.

As reported in Figure 1, some states—such as Delaware and Oklahoma—experienced much larger losses (about .35 s.d.) than other states. If such losses were allowed to become permanent, our estimates imply that students in Delaware and Oklahoma would lose 3 percent in earned income per year and, compared to other states, experience larger declines in high school graduation and college enrollment, and greater increases teen motherhood, arrests, and incarceration.

CONCLUSION

In the three decades preceding the pandemic, students in many U.S. states demonstrated large improvements in math achievement on the NAEP. Our results imply that those improvements were accompanied by a substantial improvement in later economic and social outcomes. The sizeable progress in 8th grade math achievement stands in contrast to the widespread belief that decades of K-12 school reform have been fruitless.[18] That frustration has often centered on stagnant high school achievement measures—such as the

[15] The decline in 8th grade math was 8 points, with a 2019 s.d. of 40 points. The decline in 4th grade math was 5 points, with a 2019 s.d. of 32 points.
[16] https://nces.ed.gov/programs/digest/d22/tables/dt22_203.10.asp?current=yes
[17] We assume median age, k, for those in grades K through 12 to be 6 through 18 respectively. For each k, we then calculate present value of age 28-64 earnings as $PV_k = \frac{1}{(1+r)^{(28-k)}} \sum_{age=28}^{64} \left(\frac{(1+p)}{(1+r)}\right)^{age-28} Earnings_{age}$, where p is the rate of productivity growth (assumed to .01) and r is the discount rate (assumed to be .03). $Earnings_{age}$ are the mean annual earnings of full-time, full-year workers in the 2019 ACS. We then take the mean across k.
[18] For example, see Bloomberg Philanthropies (2021) and Peterson (2016).

12th grade NAEP and the OECD's PISA exam (Blagg and Chingos, 2016).[19] However, high school graduation rates have been rising and more would-have-been drop-outs are reaching 12th grade—potentially dampening improvements in the high school measures. Even if high school achievement measures were stagnant, our results suggest that the improvements in 8th grade math achievement represented real improvements in human capital in many states.

Although we have adjusted for rising parental education and regional trends by year of birth, there may have been other unmeasured socioeconomic factors, correlated with the achievement increases, also driving higher earnings. If we have overstated the role that rising achievement played, we may also be overstating the losses from the recent declines. Nevertheless, the impacts are too large to ignore—especially in states with the largest declines. Even if the losses from permanent achievement losses were a quarter of what we estimate ($900 billion), they would still make the $190 billion in federal pandemic relief for K-12 schools seem like a worthwhile investment if students could be made whole.

[19] There are no state-level estimates of 12th grade achievement. However, in Appendix Figure 1, we show that 4th and 8th grade math achievement for U.S. students were increasing not just on the Main NAEP, but on a range of national and international assessments. High school achievement has been relatively flat, not just on the NAEP but on the international assessments as well.

REFERENCES

Barro, Robert J. (1991) "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, May, 106(2), (pp. 407– 43).

Bils, Mark and Peter J. Klenow (2000) "Does Schooling Cause Growth?" *American Economic Review*, Vol. 90, no. 5, December, (pp. 1160-1183).

Blagg, Kristin Matthew M. Chingos (2016) V*arsity Blues: Are High School Students Being Left Behind?* (Washington, DC: Urban Institute Research Report).

Bloomberg Philanthropies (2021) "Bloomberg Philanthropies Launches Historic $750 Million National Effort to Advance Public Charter Schools As Traditional District Schools Fail Children During COVID-19 Pandemic" Press Release, December 1, 2021.

Dee, Thomas S. and Brian Jacob (2011) "The impact of no Child Left Behind on student achievement" Volume30, Issue3 Summer 2011 (pp. 418-446) https://doi.org/10.1002/pam.20586

Deming DJ, Cohodes S, Jennings J, Jencks C. (2016) "School Accountability, Postsecondary Attainment and Earnings". Review of Economics and Statistics. 98 (5) :848-862.

Donohue, John J., Steven D. Levitt, (2020) "The Impact of Legalized Abortion on Crime over the Last Two Decades, *American Law and Economics Review* Volume 22, Issue 2, (pp. 241–302), https://doi.org/10.1093/aler/ahaa008

Goldhaber, Dan, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson and Douglas Staiger (Forthcoming, 2022) "The Educational Consequences of Remote Instruction during the Pandemic" *AER-Insights*.

Goldhaber, D., & Özek, U. (2019). How much should we rely on student test achievement as a measure of success? *Educational Researcher*, 48(7), 479-483.

Hanushek, E. A., & Raymond, M. E. (2005). "Does school accountability lead to improved student performance?" Journal of Policy Analysis and Management, 24, (pp. 297–327).

Hanushek, Eric A., and Ludger Woessmann (2008) "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature*, 46 (3): 607-68. DOI: 10.1257/jel.46.3.607

Krueger, Alan B. (2003) "Economic Considerations and Class Size", *The Economic Journal*, Feb., 2003, Vol. 113, No. 485, Features (Feb., 2003), pp.F34-F63.

Lovenheim, Michael F., and Alexander Willén. 2019. “The Long-Run Effects of Teacher Collective Bargaining.” *American Economic Journal: Economic Policy*, 11 (3): 292-324.

Murnane, Richard J., John B. Willett and Frank Levy (1995) “The Growing Importance of Cognitive Skills in Wage Determination” R*eview of Economics and Statistics* Vol. 77, No. 2 (May, 1995), (pp. 251-266).

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, John H. Tyler (2000) “How important are the cognitive skills of teenagers in predicting subsequent earnings?” *Journal of Policy Analysis and Management*, Vol. 19, No. 4, 547–568 https://doi.org/10.1002/1520-6688(200023)19:4%3C547::AID-PAM2%3E3.0.CO;2-%23

National Academies of Sciences, Engineering, and Medicine (2022) A Pragmatic Future for NAEP: Containing Costs and Updating Technologies. *Washington, DC: The National Academies Press*. https://doi.org/10.17226/26427.

Neal, Derek A. and William R. Johnson (1996) “The Role of Premarket Factors in Black-White Wage Differences” *Journal of Political Economy*, Vol. 104, No. 5, (pp. 869-895).

Peterson, Paul E. (2016) “Post-Regulatory School Reform” *Harvard Magazine*, September-October.

Ruggles, Stephen, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 12.0 [dataset]. Minneapolis, MN: IPUMS, 2022. https://doi.org/10.18128/D010.V12.0

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990-2022 Mathematics Assessments.