

NBER WORKING PAPER SERIES

HAVE INCOME-BASED ACHIEVEMENT GAPS WIDENED OR NARROWED?

Shirin A. Hashim
Thomas J. Kane
Thomas Kelley-Kemple
Mary E. Laski
Douglas O. Staiger

Working Paper 27714
<http://www.nber.org/papers/w27714>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2020, Revised March 2022

Shirin Hashim, Mary Laski and Thomas Kelley-Kemple received support from the Walton Family Foundation for their work on this project. Thomas Kelley-Kemple also received support under pre-doctoral training grant R305B150010-16 from the Institute of Education Sciences at the U.S. Department of Education. We thank Aimee Chen, Janet Currie, David Deming, John Friedman, David Grissmer, Eric Hanushek, Caroline Hoxby, Dan Koretz, Eric Nielsen, Paul Peterson, Steve Raudenbush, Sean Reardon, Sarah Reber, Jesse Rothstein, Sarah Turner, Jim Wyckoff and seminar participants at the University of Virginia, Harvard Graduate School of Education, 2020 NBER Summer Institute, Harvard Opportunity Insights, and the 2020 Association for Public Policy Analysis and Management annual meeting for helpful comments and suggestions. We thank Blake Heller for help preparing data for analysis. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w27714.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Shirin A. Hashim, Thomas J. Kane, Thomas Kelley-Kemple, Mary E. Laski, and Douglas O. Staiger. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Have Income-Based Achievement Gaps Widened or Narrowed?

Shirin A. Hashim, Thomas J. Kane, Thomas Kelley-Kemple, Mary E. Laski, and Douglas O. Staiger

NBER Working Paper No. 27714

August 2020, Revised March 2022

JEL No. I21,I22,I24

ABSTRACT

Over the past 30 years, rising income inequality and income-based residential segregation have threatened to widen income-based achievement gaps, even as school accountability and school finance reform efforts have attempted to narrow them. Yet, no national dataset measures both parental income and achievement in a consistent way for individual students over time. We take two alternative approaches to inferring income-based achievement gaps: First, we reconstruct the student-level relationship using school-level estimates of means and variances of achievement and income. Second, we combine estimates of mean income by race, mother's education, urbanicity and state with mean achievement for the corresponding subgroups on a national assessment. Using both methods, we find that income-based achievement gaps in 4th and 8th grade narrowed between 1992 and 2015—while math scores rose at all income levels.

Shirin A. Hashim
13 Appian Way
Harvard University
Cambridge, MA 02138
shirin_hashim@g.harvard.edu

Mary E. Laski
13 Appian Way
Harvard University
Cambridge, MA 02138
marylaski@g.harvard.edu

Thomas J. Kane
Harvard Graduate School of Education
Center for Education Policy Research
50 Church St., 4th Floor
Cambridge, MA 02138
and NBER
kaneto@gse.harvard.edu

Douglas O. Staiger
Dartmouth College
Department of Economics
HB6106, 301 Rockefeller Hall
Hanover, NH 03755-3514
and NBER
douglas.staiger@dartmouth.edu

Thomas Kelley-Kemple
13 Appian Way
Harvard University
Cambridge, MA 02138
tkelleykemple@g.harvard.edu

I. Introduction

In 1848, Horace Mann famously described public education as “the great equalizer of the conditions of men” and the “balance wheel of the social machinery.” That balancing function has been strained in recent decades, as income inequality and income-based residential segregation have increased.

Unfortunately, we know remarkably little about how income-based achievement gaps have changed over time. In the United States, there is no nationally representative data source combining a consistent measure of academic achievement with the household income of individual students.¹

In a widely cited 2011 paper, Sean Reardon pooled data from twelve nationally representative surveys to estimate the relationship between achievement and income by birth cohort. After adjusting for the different assessments and income measures used in each survey, Reardon (2011) concluded that the difference in mean achievement at the 90th and 10th percentiles of family income had risen by 40 to 50 percent for those born between 1974 and 2001.

We take two alternative approaches, capitalizing on the consistent assessment scale provided by the National Assessment of Educational Progress (NAEP). First, we supplement the NAEP data with estimates of the mean and variance of income in the neighborhoods surrounding each sampled school. We use school-level means for achievement and income to estimate the between-school slope and school-level variances in achievement and income to estimate the within-school slope.

¹ Given that few 4th and 8th grade students have direct knowledge of their parents' incomes, the preferred approach to measuring family income would be to include a parent survey. Yet, Title III Section 303(c)3(B) of the Education Sciences Reform Act presents a major obstacle to adding a parent survey to the National Assessment of Educational Progress, by prohibiting NAEP contractors from maintaining “any system of records containing a student's name, birth information, Social Security number, or parents' name or names, or any other personally identifiable information.”

(Intuitively, the schools with more unequal incomes should also have proportionally more variance in achievement, and the slope of that relationship should be equivalent to the squared value of the within-school slope). We then reconstruct the student-level coefficient using a weighted average of the within- and between-school coefficients with weights based on the intra-class correlation in income by school.

Because we rely on census data to estimate the mean and variance of parental incomes by school, a decline in the reliability of our income measure could lead to a flattening of the estimated relationship between income and achievement due to attenuation bias. Yet, when we investigate the correlation between our imputed income measure and the publicly reported measures of free and reduced-price lunch receipt, the correlation is stable over time. Moreover, whether we adjust analytically for measurement error using the sample size of households in each neighborhood or instrument the income measures using lagged values from earlier decennial censuses, our findings point to a narrowing of income-based achievement gaps.

In our second set of estimates, we use the March Current Population Survey (CPS) to impute household income by race/ethnicity, mother's education, urbanicity and state—all characteristics that are measured at the student level in the NAEP. We then investigate how the gaps in achievement by imputed income changed over time. With student-level imputed income, we can re-estimate the between-school and within-school coefficients—as a check of our findings with school-level aggregates.

In contrast to Reardon (2011), we see no evidence of widening of achievement gaps in 4th grade or 8th grade math or reading using either method. In fact, we see achievement gaps *narrowing*. For example, between 1992 and 2015, we estimate that the achievement of students at the 10th percentile of income improved substantially: a full standard deviation in 4th grade math

and .7 standard deviations in 8th grade math. Although achievement also rose for higher income families, the achievement gap between those at the 90th and 10th income percentile *closed* by .4 and .3 standard deviations in 4th and 8th grade math respectively.²

During the period when Reardon concluded income-based achievement gaps were expanding, the Black-White achievement gap in the NAEP was moving in the opposite direction, narrowing by .3 and .2 standard deviations in 4th grade and 8th grade math respectively between 1992 and 2015. Given racial differences in family incomes, we show that Reardon's findings would have required the Black-White gap in test scores *conditional on income* to change sign. In other words, to reconcile a widening income-gap and a closing racial gap, Black students would have had to have higher achievement than White students of the same income in the later years. In contrast, our finding of narrowing of income-based achievement gaps is consistent with no change in the income-adjusted racial gap.

In the following discussion, we briefly summarize the prior literature, describe our methods, and present our results.

II. Prior Literature

In a widely cited paper, Reardon (2011) compared estimates from 12 different surveys, each using a different measure of achievement and income, to infer whether gaps were widening over time. He reports that reading and math achievement gaps at the 90th and 10th family income percentiles grew by 40 to 50 percent among cohorts born between 1974 and 2001.

To compare the different surveys, Reardon was required to make a number of strong assumptions. First, because he pools surveys using different student assessments, Reardon assumes

² Although both of our estimators focus on public schools, we present evidence in Appendix A that Reardon's results cannot be explained by differential trends in public and private school achievement.

that income gaps are the same by type of test and by age (whether a picture vocabulary test administered to pre-school children or the ASVAB test administered to high school age students). Second, each of the surveys uses a different income measure. For example, depending on the survey, the number of income categories respondents could choose from ranged from five to fifteen categories. The surveys prior to 1975 often included student-reported family income; all the surveys for birth cohorts after 1975 rely on parent-reported income.³ Although he adjusts for the reliability of the income measures, he assumes that the reliability of parent-reported income and student-reported income (by student age group) was the same over time and across surveys. While defensible, each of these adjustments is necessarily somewhat arbitrary and adds uncertainty to the estimated trend, which is not reflected in the standard errors.

To avoid having to adjust for differences in achievement and income measures, Nielsen (2019) re-analyzed two of the datasets included in Reardon's analysis. The National Longitudinal Survey of Youth in 1979 and 1997 used similar measures of parent-reported income and similar assessments, thus reducing the assumptions required for establishing comparability. Using fully ordinal methods (that do not rely on Reardon's assumption of a consistent linear scaling across the various achievement measures), Nielsen finds income-based gaps narrowing. When he uses Reardon's methods and a similar definition of income, he finds that gaps were stable. However, the cohorts included in Nielsen's analysis (those born between 1962 and 1982) only partially overlap with the birth cohorts for whom Reardon sees income-based gaps widening (1975 to 2001).

In a recent paper, Hanushek, Peterson, Talpey, and Woessmann (2020) use achievement data from national and international tests (PISA, TIMMS, and the NAEP) to measure gaps in

³ For the birth cohorts for whom Reardon finds widening gaps in achievement by income, the income measures are parent reported. Thus, it seems unlikely that the finding is driven by the shift from student-reported income measures (which tend to be less reliable) to parent-reported measures.

achievement by SES. After constructing an SES index using student-reported parental education and household possessions, Hanushek and colleagues find no change in reading or math achievement gaps between top and bottom SES quartiles.

However, because they are measuring SES-based gaps, the findings in Hanushek et al. (2020) are not necessarily inconsistent with Reardon's findings. Reardon also reported stable differences in achievement by parental education (a main factor in the Hanushek et al. SES index). Moreover, as reported in Appendix B, stable gaps in achievement by SES are consistent with narrowing gaps by family income, because the income differences by SES have grown over time.

Especially during the last three decades, U.S. policymakers have launched a series of policies meant to raise achievement in low-scoring and low-income schools. For instance, Dee and Jacob (2011) studied the effect of school accountability policies, which were adopted in 30 states between 1990 and 2001 and then expanded nationwide with the No Child Left Behind Act in 2002. The authors find that school accountability policies had positive effects on average achievement in 4th and 8th grade math, and weaker effects on 4th grade reading. The estimated effects were larger for Black and Hispanic students and students eligible for free and reduced-price lunch, especially in 4th and 8th grade math.

Lafortune, Rothstein and Schanzenbach (2018) studied the impact of 64 school finance reform events (SFRs) across 26 states between 1990 and 2011. They find that SFRs led to narrowing of gaps in expenditures per pupil and achievement between high and low-income school districts. Jackson, Johnson and Persico (2016) find that school finance reforms led to increased school spending and higher wages for lower income children.

The birth cohorts within Reardon's window of widening achievement gaps would have started first grade between 1980 and 2006 (when they would have been 5 years old) and would have been

enrolled in 8th grade between 1988 and 2014 (when they were 13 years old). Over that time period, there were important expansions in non-monetary benefits for low-income children.⁴ For example, Currie and Duque (2019) review the literature on the impact of the Medicaid expansions, which occurred in the 1980's and 1990's, on children's health and mortality. Currie and Schwandt (2016) find that expansions in public health insurance coverage dramatically reduced mortality among poor children. A 2019 consensus report from the National Academy of Sciences, Engineering, and Medicine concluded that the Earned Income Tax Credit (EITC) Program also improved children's health outcomes. Although the EITC was first enacted in 1975, it was sharply expanded in 1986, 1990 and 1993. Moreover, real spending on the Food Stamp program increased by 32 percent between 1990 and 1996.⁵ Bailey et al. (2020) found that the availability of Food Stamps before age 5 led to an increase in educational attainment as adults. Thus, Reardon's results would imply that gaps in achievement scores were widening while gaps in children's health and other outcomes were narrowing due to expansions in other programs targeted at low-income families.

III. Empirical Strategy

We employ two different strategies to estimate the student-level relationship between income and achievement. In the first, we use school-level aggregates to estimate the student-level relationship between income and achievement. In the second, we use the March Current Population Survey in each year to impute household income by students' race/ethnicity, mother's education, state of residence, and an indicator for urbanicity—all student-level traits available in the NAEP. We briefly describe each of these two approaches below.

⁴ We thank Janet Currie for pointing to this literature on the impacts of the Medicaid expansions.

⁵ <https://fns-prod.azureedge.net/sites/default/files/resource-files/SNAPsummary-1.pdf>

A. *Using School Level Aggregates to Estimate the Student-Level Parameters*

Our goal is to estimate the student-level relationship between achievement and income:

$$(1) \quad \text{Score}_{ij} = \beta_0 + \beta_1 Y_{ij} + v_{ij}$$

where Score_{ij} is a student's scaled test score, Y_{ij} is a measure of log household income and i and j are subscripts for the student and the student's school, respectively. (Obviously, β_1 is a descriptive relationship, not a causal one).

It can be shown (Raudenbush and Bryk, 2002, pp. 136-137) that the OLS estimator for β_1 is a weighted average of the between-school coefficient of mean income on mean achievement ($\hat{\beta}_{between}$) and the within-school relationship between an individual student's household income and achievement ($\hat{\beta}_{within}$):

$$(2) \quad \hat{\beta}_1^{OLS} = \hat{\beta}_{between} * ICC_Y + \hat{\beta}_{within} * (1 - ICC_Y)$$

where ICC_Y is the percent of the total variance in log household income that is attributable to school-level differences in mean log income ($ICC_Y = \frac{\sigma_{\bar{Y}_j}^2}{\sigma_{Y_{ij}}^2}$).⁶ With a value between 0 and 1, ICC_Y is a measure of the degree of income segregation by school. That is, if all the variation in income was by school (every student in a given school had the same income) and different schools had different incomes, then the student-level relationship between achievement and income would be equal to the between-school relationship.

If the parameters in equations (1) and (2) were estimated with student-level data, the two equations would yield numerically identical estimates of β_1 . However, our estimation of equation

⁶ The statistic ICC_Y is the empirical analog of the intra-class correlation in log income by school. Bryk and Raudenbush (1992) refer to the statistic, ICC_Y , as "eta squared". In contrast to the traditional "intra-class correlation", or ICC, which uses an estimate of the variance of the true school mean incomes in the numerator, μ_{Y_j} , the statistic, ICC_Y , uses the variance in the sample means, \bar{Y}_j (which has more variance because \bar{Y}_j is equal to μ_{Y_j} plus estimation error). In a setting with a large sample (or the entire population) for each school, the two measures are approximately equivalent.

(2) differs from the ideal in two ways. First, we have school-level data (not student-level data) on the means and variances of income and test scores, which necessitates additional assumptions to estimate $\hat{\beta}_{within}$. Second, our school-level income data come from a different sample than the test score data, which introduces measurement error and attenuation bias. We discuss how we address each of these issues in turn.

Estimation With School-Level Data.—With school-level measures of mean achievement and mean log income, \overline{Score}_j and \bar{Y}_j , we can estimate the between-school relationship directly:

$$(3) \quad \overline{Score}_j = \beta_0 + \beta_{between} \bar{Y}_j + \varepsilon_j$$

We can also estimate the within-school slope using school-level data on the variance in achievement and income, $\hat{\sigma}_{score_j}^2$ and $\hat{\sigma}_{Y_j}^2$:

$$(4) \quad \hat{\sigma}_{score_j}^2 = \sigma_v^2 + \beta_{within}^2 \hat{\sigma}_{Y_j}^2 + \vartheta_j$$

where σ_v^2 is the variance of the residual from the within-school regression. The basic intuition behind equation (4) is that schools with higher income variance (i.e., more income inequality) should have proportionally higher variance in achievement—with the proportion determined by the square of the within-school slope coefficient.

For OLS estimation of equation (4) to yield consistent estimates of β_{within}^2 requires the additional assumption that $\hat{\sigma}_{Y_j}^2$ and ϑ_j are uncorrelated. In other words, we assume that the variance of income within a school is not correlated with other factors affecting the variance of achievement. If schools with more heterogeneity in income have more (less) heterogeneity in achievement conditional on income, our estimate of β_{within}^2 will be biased upward (downward). More

importantly for our analysis, if the correlation between $\hat{\sigma}_{Y_j}^2$ and ϑ_j is changing over time, our estimates of the trend in β_{within} will be biased.

As a concrete example, suppose that conditional on family income, student achievement was also affected by mother's education. We are assuming that the variance of education (and its impact on achievement) within income groups is the same in schools with high- and low-income variance and does not change over time.

Another way in which OLS estimation of equation (4) could be biased is if β_{within} varied across schools – i.e., if some schools were better able to reduce income-based achievement gaps than others. In equation (4), we express β_{within} as a scalar, assuming that the within-school slope is the same in all schools. However, equation (4) could be generalized to allow the slope to be a random coefficient, varying by school. If the within-school slope did vary by school, but was independent of log income, then the coefficient on $\hat{\sigma}_{Y_j}^2$ in equation (4) would be interpreted as the sum of the expected value of β_{within} squared and the variance in β_{within} (that is, $E[\beta_{within}]^2 + \sigma_{\beta_{within}}^2$). So long as the variance in β_{within} ($\sigma_{\beta_{within}}^2$) is constant or rising, then any increase in the expected value of β_{within} should lead to an increase in the coefficient on $\hat{\sigma}_{Y_j}^2$ in equation (4). It is only if the variance in within-school slopes is declining that we might not detect an increase in the within-school slope using our method.

Finally, in estimating equations (3) and (4), we weight by the number of tested students in each school. We also allow the errors to be clustered by the geographic primary sampling unit used by NAEP. Given that the error terms, ε_j and ϑ_j , could be related, we estimate equations (3) and (4) as seemingly unrelated regressions (SURE). Finally, as described in equation (2), the student-level slope is a nonlinear function of the coefficients in equations (3) and (4), $\beta_{between}$ and β_{within}^2 .

Thus, we use the delta method to estimate a standard error for our estimate of the student-level slope ($\hat{\beta}_1^{OLS}$).

Accounting for Measurement Error in School-Level Income Data.—The school-level income data are drawn from the census, rather than from the sample of students included in the NAEP, which introduces two main sources of measurement error. First, the census tract boundaries often do not correspond with the school catchment area boundaries. Since this measurement error is not necessarily classical, it is unclear how this will bias our estimates or whether the direction of this bias is likely to change over time. To explore how well nearby tracts proxy for school characteristics and whether this changes over time, we use estimates of school demographics (race, free and reduced-price lunch eligibility) derived from nearby census tracts to predict administrative data on school characteristics from the federal Common Core of Data. If census-derived estimates of school demographics are stable predictors of actual school demographics over time, then we have greater confidence in the census-derived estimates of school income.

A second source of measurement error in our school-level income data arises because the census income data are based on a sample of households in each tract, which introduces potential measurement error and resulting attenuation bias. Our census income estimates come from the 1990 and 2000 decennial censuses. The data from 2009 and onward are derived from the American Community Survey (ACS). Importantly, the proportion of total households that were sampled declined over time: 20 percent of the population were asked the income questions in 1990, 16 percent in 2000, and only 5 percent beginning in 2009 (pooling 5 years of 1 percent samples in the ACS data). As the sample size declines over time, the measurement error in school income will rise leading to greater attenuation of the coefficients in estimating equations (3) and (4) by OLS (and thus an apparent decline in the relationship between scores and income over time).

Thus, we take two approaches to correcting estimates of equations (3) and (4) for attenuation bias. First, we calculate analytic estimates of the share of the total variance in \bar{Y}_j and $\hat{\sigma}_{\bar{Y}_j}^2$ that is due to estimation error and use these to dis-attenuate OLS estimates of equations (3) and (4). We calculate the variance of the estimation error in \bar{Y}_j and $\hat{\sigma}_{\bar{Y}_j}^2$ based on sample sizes in each census and standard formulas for the estimation error for a sample mean and variance, and then correct for the implied attenuation bias arising from this amount of estimation error (see Appendix C for details). Second, we estimate equations (3) and (4) by 2SLS, using estimates of \bar{Y}_j and $\hat{\sigma}_{\bar{Y}_j}^2$ from an earlier census year as instruments for the corresponding estimates from the current year (e.g., using 1980 census estimates as instruments for years based on 1990 census estimates, the 1990 census estimates as instruments for the years based on the 2000 estimates, and so on). Using one noisy estimate to instrument for another is a standard correction for attenuation bias: estimates from a prior census are valid instruments because they are based on an independent sample and therefore uncorrelated with the measurement error in the current estimates.

B. Using Imputed Household Income by Student Traits

For our second set of estimates, we impute income using student-level traits collected in the NAEP (race/ethnicity, mother's education, urbanicity, and state.)⁷ We estimate the mean household income for youth with those same traits in the Annual Social and Economic Supplement to the Current Population Survey. Using imputed household income for each year, we estimate the student-level slope, as well as the between- and within-school slopes. (We can estimate all three

⁷ Students participating in the 8th grade NAEP assessments were asked to report education levels for their mother as well as their father. (The survey measures of parents' education were not consistently available for the 4th grade students in NAEP). However, the NAEP did not consistently ask whether either parent lives in the same household with the student. Therefore, we assume that students in the NAEP lived in households including their mother when imputing household income. (We do not use information about the father's education, given that we cannot know whether the father is present).

directly since we have imputed household income at the student-level).

Specifically, we estimate the following relationship for all households in the Current Population Survey with dependent children between age 5 and 18 with a mother present:

$$(5) \quad Y_{it} = \sum_{\text{race/eth}} \sum_{\text{moth ed}} \pi_{\text{race,moth ed } t} + \gamma_{\text{urban}, t} + \delta_{\text{state}, t} + \varphi_{it}$$

where Y_{it} is a measure of log household income, $\pi_{\text{race,moth ed } t}$ represents a full set of interactions between five categories of mother's education categories and six categories for race/ethnicity, $\gamma_{\text{urban}, t}$ reflects three categories of urbanicity, and $\delta_{\text{state}, t}$ represents fixed effects by state.⁸ We run regressions separately for each year, allowing the coefficients to change over time. As reported in Appendix Table D.1, the R^2 for each year was largely unchanged at roughly .25 in each year.

Using the parameter estimates from equation (5), we generate a predicted household income for each 8th grade respondent with non-missing mother's education, \hat{Y}_{it} . Since we are using imputed income, \hat{Y}_{it} , to substitute for actual household income, we would rewrite equation (1) as:

$$(1') \quad \text{Score}_{it} = \beta_{0t} + \beta_{1t} \hat{Y}_{it} + (\beta_{1t} \hat{\varphi}_{it} + v_{it})$$

where $Y_{it} = \hat{Y}_{it} + \hat{\varphi}_{it}$. Note that as long as \hat{Y}_{it} is estimated by OLS, then \hat{Y}_{it} and $\hat{\varphi}_{it}$ will be orthogonal by construction. In other words, even though imputed household income is an imperfect measure of true household income, our estimate of β_{1t} in equation (1') will not be biased due to measurement error.

Many of the variables being used for the imputation—mother's education, race, state, urbanicity—are likely to have their own direct effects on achievement beyond their influence on family income (that is, \hat{Y}_{it} will be correlated with the error term from equation (1), v_{it}). We are not assuming that the coefficient on imputed income is the same as the coefficient on income,

⁸ We infer that a mother is present if there is a female head of household, a female spouse of the head of household, and/or a female unmarried partner. If there are multiple potential mothers present, we use the maximum of their education.

which we do not observe. Rather, we are merely trying to learn about the trend in the achievement-income relationship. For the latter, we must assume that the correlation between \hat{Y}_{it} and v_{it} is constant over time and that any change in $\hat{\beta}_{1t}$ is due to the effect of income and not to changes in the direct effects of the variables used for the imputation. (Thus, we also use different subsets of household traits—race, mother’s education, state and urbanicity—as a robustness test, in case the direct effect one such trait is driving our results).

Just as with the usual 2SLS estimate, equation (1’) will estimate a LATE that may differ from that estimated by OLS. Recall from equation (2) that $\hat{\beta}_1^{OLS}$ is a weighted average of the within-school and between-school slopes, where the weight on the between-school slope is ICC_Y (the intra-class correlation in log income by school). Similarly, we can restate the estimate from equation (1’) using imputed log income (call this $\hat{\beta}_1^Y$) as:

$$(2') \quad \hat{\beta}_1^Y = \hat{\beta}_{between}^Y * ICC_{\hat{Y}} + \hat{\beta}_{within}^Y * (1 - ICC_{\hat{Y}})$$

If we assume that \bar{Y}_i and $\bar{\varphi}_i$ (the school-level means of \hat{Y}_{it} and $\hat{\varphi}_{it}$) are orthogonal (an additional assumption not guaranteed by OLS), then it can be shown that $\hat{\beta}_{between}^Y = \hat{\beta}_{between}$, $\hat{\beta}_{within}^Y = \hat{\beta}_{within}$ even when $ICC_Y \neq ICC_{\hat{Y}}$. In other words, if the error in equation (5) is uncorrelated with the proxy measure of income at both the individual level (true by construction) and at the school level (an assumption), then the within and between slopes are the same for the true income data as for the imputed, proxy income measures. In this case, our estimate of the overall relationship between scores and log income using actual income (equation (2)) differs from our estimate using imputed income (equation (2’)) because the two estimates put different weight on the between and within slopes (since in general $ICC_Y \neq ICC_{\hat{Y}}$).

Therefore, although we estimate equation (1’) with student-level data, we also estimate the between-school and within-school coefficients and the ICC (from equation (2)) using imputed

student level data. We then evaluate whether any difference in our estimates of β_{1t} between our two methods are due to differences in the within- and between-school estimates (which should show similar trends) or simply differences in the weighting placed on within versus between estimates due to differences in the ICC (which are expected).

IV. Data

Known as “the Nation’s Report Card,” the main NAEP assessment collects data on a representative sample of schools in each participating state (roughly 4,800 schools per year per grade level during the period we study) and a random sample of roughly 30 students in each school. Importantly, the 4th and 8th grade NAEP assessments are designed to provide scores on a consistent scale over time.⁹ Prior to 2002, states were able to opt out of participating, and the number of participating states ranged from 38 to 46, depending upon the subject and grade. Since 2003, participation has been mandatory and the NAEP has been administered every other year, in both reading and mathematics. Balancing the twin goals of maintaining a consistent sample of states and remaining as close to nationally representative as possible, we included states that missed at most one year of assessment scores in a given grade-subject combination. Thus, in 4th grade math and reading, that left us with 45 and 42 states respectively, and 45 and 46 states in 8th grade math and reading respectively (representing approximately 88-94 percent of students in each grade/subject in 2015).

To add data on household income for public schools, we use each school’s geographical coordinates from the Common Core of Data (CCD). We rank the nearby census block groups by

⁹ For more information on how NAEP maintains the comparability of scores over time, see <https://nces.ed.gov/nationsreportcard/tdw/analysis/trans.aspx>.

their distance from each school.¹⁰ Using the number of students enrolled in grades 1-4 and 5-8 reported for each block group, we identify the set of closest block groups which would be sufficient to “fill up” each school, based on the school’s enrollment reported in the CCD. We assume that the population in each block group is representative of its census tract and sum the number of families with children in various income bins across the set of tracts associated with each school. To calculate the mean and variance of log income for each school, we fit each school’s income bins to a log-normal distribution, using a method outlined by vonHippel, Scarpino, and Drown (2016). (See Appendix D for more details). We repeat the process using the 1990 and 2000 decennial censuses (and, after 2009, the ACS). For intervening years, we use a linear interpolation of the mean and variance of income from 1990, 2000, and 2009.

Our estimates of the mean and variance in family income for individual public schools will suffer from three sources of measurement error: first, census block group boundaries may not coincide with actual school attendance zones; second, the census and the ACS estimates are based on random samples of households in each census tract (not the population); third, not all children in a neighborhood will be attending public schools. According to the National Center for Education Statistics (2020), 13.1 percent of students in grades pre-K through 8 attended private schools in 1990 (around the beginning of our period). The percentage had declined slightly to 10.8 percent by 2016.

Nevertheless, the method performs reasonably well in replicating administrative data on race and income from the Common Core of Data (CCD). To check, we compare our census-based estimates of poverty rates and race/ethnicity to the proportion of students receiving free or reduced-

¹⁰ Overlaying school attendance zones and block group boundaries for a set of 21 of the largest 22 districts, Saporito and Sohoni (2007) found that 53 percent of census block groups in 2000 lay entirely within a school boundary in 2007.

price lunch (FRPL) and each race/ethnic group in the CCD. Those results are reported in Table 1. Unfortunately, the data on students receiving free and reduced-price lunch was missing for more than half of schools in 1990 (52 percent). In 2000, the correlation between the poverty rates in nearby block-groups and the school's reported percentage of students receiving free and reduced-price lunch was .715. By 2015, the correlation was roughly the same, .690.¹¹ The method performs better for approximating a school's race-ethnicity; the correlation between our estimate of the percentage of students who are Black in each school and the CCD report was roughly .9 in all years between 1990 and 2015. Although the census-based measure of school income is subject to error, there is no evidence that the reliability of our estimates has diminished over time.

V. Using School-Level Data to Estimate Student-Level Parameters

Before presenting parameter estimates, we start with a graphical summary of the between-school and within-school relationships in various years. Figure 1 reports the results of a locally weighted polynomial regression of school mean achievement on mean log income for 4th and 8th grade math and reading. The solid lines portray the relationship for the first year (1992 for 4th grade math and reading and 8th grade math; 1998 for 8th grade reading) and the last year (2015) in our analyses; the dashed lines report the relationships for the years between 1992 and 2015. Upon inspection, the relationship is roughly linear in log income. Moreover, in both subjects—but especially in math—there has been a noticeable increase in the intercept and a decline in the slope. At least in terms of the differences in mean achievement between schools, there is no evidence of

¹¹ The imputation performs best in suburban areas, with an R^2 of approximately .75 with respect to FRPL status, but R^2 between .6 and .65 in the urban and rural areas as well as towns. The correlations for primary schools (those with 4th grade students) and middle schools (which we define as those with 8th grade students) were also similar and consistent over time.

a sharp steepening of the relationship between mean achievement and mean income between schools.

Figure 2 reports the trend in the intra-class correlation in log income for schools containing a 4th grade and for schools including an 8th grade. Likely reflecting the fact that attendance zones for elementary schools are smaller than for middle schools, the ICC for 4th grade was somewhat higher than for 8th grade in 1992 (.19 versus .16). However, both have increased by roughly .06 points over time, implying a roughly 30 percent increase in income-based school segregation by that metric.

Figure 3 includes plots of the relationship between the school-level variance in achievement and within-school variance in log income the four grade/subject combinations in the first year and the last year we see them. Sample-based estimates of variances are inherently noisy (especially for variances), so to summarize the central tendency, we have calculated the average variance in achievement and the average variance in income for 20 equal-sized bins arranged by variance in log income. We have plotted the mean variance in achievement for each bin and then fitted a line. The relationship does appear to be roughly linear, with an increase in achievement variance being approximately proportional to the rise in log income variance. However, there is again no evidence of a steepening slope for any grade-subject combination. If anything, the slopes appear to be declining over time.

Our estimate of the within-school slope will be imprecisely estimated, given that we are regressing a sample variance in achievement on an estimate of income variance. However, equation (1) lets us bound the magnitude of the increase in the within-school slope that would be required to be consistent with Reardon's finding. If there has been no change in the between-school slope and the ICC is roughly .2, a 40 percent increase in the overall slope that Reardon (2011)

infers would have required a 50 percent increase in the within-school slope. Recall that the within-school slope is the square of the variance-on-variance relationship. So, a 50 percent increase in the within-school slope would require more than doubling of the slopes depicted in Figure 3 ($1.5^2=2.25$). Even if we cannot estimate it precisely, we see no evidence of a 50 percent increase in the average within-school slope in any grade/subject combination.

In Table 2, we report our estimates of the between-school and within-school slope parameters and then use the estimated ICC to reconstruct the student-level relationship for the first year and last year available for each grade/subject. (For estimates for all available years, see Appendix Tables E.1 through E.4). We start by discussing 4th grade math. As reported in column (1), our point estimate of the between-school slope declined from .867 to .657 between 1992 and 2015, approximately a 25 percent decline. In column (2) we report estimates of the within-school slope parameter. The within-school coefficient also declines from .454 to .179 (also statistically significant). As reported in column (3), the intra-class correlation in log income, which is a measure of income-based segregation, rises from .19 to .26.¹² In column (4), we combine the between-school slope and the within-school slope using the ICC to generate an estimate of the student-level slope parameter. Our results imply that the student-level slope declined from .533 to .302 between 1992 and 2015.¹³ Thus, we estimate that a 10 percentage point difference in family income was associated with a .05 standard deviation higher score in 4th grade math in 1992; by 2015, that had declined to .03 standard deviations.¹⁴

¹² An increase in the ICC would have led to a small increase in the student-level slope if the between- and within-school coefficients had both remained constant (because the between-school slope is larger than the within-school slope.) However, the rise in ICC was not nearly large enough to offset the declining between-school and within-school slopes.

¹³ The change in the between-school, within-school and student-level slopes are all statistically significant.

¹⁴ The standard deviation in 4th grade math was 32 points in the 1992 NAEP.

In the remaining columns of Table 2, we report the implied change in mean achievement at the 10th and 90th percentile of log household income. (Because the NAEP does not have student level data on income, we are estimating these by plugging in the relevant percentiles of household income from March CPS data. See Appendix Table E.5). In column (5), we report the estimated mean achievement at the 10th percentile of family income in each year. We estimate that there has been a .998 standard deviation increase in the mean 4th grade math NAEP score for students from families who had incomes that placed them in 10th percentile in 1992, from -.645 to .353. In column (6), we report the estimated mean achievement at the 90th percentile of family income in each year. Achievement increased for high-income students as well, by .564 standard deviations, from .480 to 1.045. Thus, between 1992 and 2015, we estimate that the gap in 4th grade math achievement between the 90th and 10th income percentiles had shrunk by .43 standard deviations.

The remaining panels in Table 2 report results for 8th grade math, 4th grade reading, and 8th grade reading. The student-level slopes declined in all four grade/subject combinations; we estimate that the gaps for those at the 90th and 10th percentiles of income closed by .32 standard deviation in 8th grade math, .80 standard deviations in 4th grade reading and .35 standard deviations in 8th grade reading. One anomaly worth noting is that, in 4th grade reading, we estimated a small negative slope in the variance-on-variance regression in the two latter years—2013 and 2015. (Because the estimated slope was negative, the square root was not identified). We interpret a negative slope at the boundary constraint of zero. We do not literally believe that there was no relationship between income and achievement within school, but we do take these results as evidence that the within-school income gaps declined over time.

For the early birth cohorts, our estimates of the income-based achievement gaps are similar in magnitude to those reported by Reardon (2011). For instance, Reardon reports that size of the

achievement gap at the 10th and 90th percentile of income was 1 standard deviation for the 1976 birth cohort. In Table 2, we estimate the 90-10 gap to have been 1.13 and 1.20 in 4th grade math and reading respectively for the 1983 birth cohort (the 9-year-olds in 1992), 1.31 in 8th grade math for the 1979 birth cohort, and 1.02 standard deviations in 8th grade reading for the 1985 birth cohort. However, between the 1976 and 2001 birth cohorts, Reardon estimates that the math achievement gap at the 90th and 10th percentiles grew by 40 percent, while our estimates suggest that 90-10 gap declined substantially in all four grades/subjects.

Adjusting for Measurement Error.—As noted in the methodology section, the number of households we use to estimate neighborhood income declined over time, from 1 in 5 households in 1990 to 1 in 6 households in 2000 to roughly 1 in 20 households in the 5-year sample from the ACS. An increase in attenuation bias could contribute to a narrowing estimated gap with respect to income. In this section, we estimate the magnitude of the potential bias.

First, we use estimates of the mean and variance in neighborhood income during the prior decennial census to instrument for current year values. In columns (1) through (4) of Table 3, we report the student-level coefficients implied by the IV.¹⁵ To facilitate comparison to the OLS estimates, we also report the student-level coefficients from Table 2 in columns (5) through (8). The IV estimates are consistently larger in magnitude than the estimates from OLS. (We would expect such a pattern if there were any measurement error in the measures). However, both the IV and the OLS estimates imply a declining slope of achievement with respect to income. For example, in 4th grade and 8th grade math, the IV estimates decline by 40 percent and 13 percent respectively between 1992 and 2015, while the OLS estimates decline by 43 and 31 percent.

¹⁵ To generate the results in Table 3, we estimated separate specifications for the between-school and within-school coefficients, using lagged values of the mean and variance in income respectively. In Table 3, we report the weighted combination using the ICC. See Appendix F for the between-school and within-school IV estimates by grade/subject).

In Figure 4, we present three sets of estimates for each of the four grade/subject combinations: (i) our original estimates of the student-level coefficient; (ii) the IV estimates; and (iii) the original estimates after applying the analytic measurement error correction described in the methods section (see Appendix C for details). In all four grades and subjects, we see declining slopes using all three approaches. Although the magnitude of the decline in slope is slightly smaller with the IV and analytic measurement error adjustments, we do not see evidence that measurement error is driving our results.

VI. Using Student Traits to Impute Income

Although the NAEP does not collect information on parental income, it does collect information on a number of student and family characteristics that are correlated with income, such as race/ethnicity, mother's education, urbanicity, and state of residence, which we use to impute household income. In Table 4, we report the coefficients on imputed income using each student trait separately (race/ethnicity, mother's education, urbanicity, and state) and then combined (using interactions between race/ethnicity and mother's education) for 8th grade math and reading. Analogous to IV, we are essentially scaling the differentials in achievement associated with each of these traits by the magnitude of the associated family income differential from the March CPS. The trends over time should be consistent with the trends in underlying income effects as long as the *direct* effects of race, mother's education, urbanicity, and state are constant.

The implied slope with respect to log income for 8th grade math based on the race differentials declined by 17 percent between 1992 and 2015, from 1.435 to 1.186. Similarly, using mother's education to impute income, the implied slope with respect to log income declined by 10 percent. Using state and urbanicity, the implied slope declined by 36 percent, from .805 to .517—largely

because many low-income states closed the gap with respect to higher income states. When we combine all the measures, the implied slope declined by 14 percent, from .938 to .804.

In the top panel of Figure 5, we plot the mean standardized 8th grade math score (from the NAEP) and the mean family income (from the March CPS) for each subgroup by race/ethnicity and mother's education by urbanicity by state. For illustration, we call out two specific points: the mean income and achievement of children of Hispanic high school dropouts from rural New Mexico and the mean income and achievement of children of White college graduates from suburban New York. We then estimate the relationship between mean achievement and mean income across those subgroups. We report the same relationship in 1992 and 2015. Among 8th grade students, mean math achievement rose at all income levels, but particularly for low-income groups, with the implied relationship between achievement and income flattening.

In the bottom panel of Figure 5, we report a similar relationship for 8th grade reading in 1998 (the first year of the state sample in 8th grade reading) and in 2015. Our results imply that the increase in 8th grade reading achievement was concentrated in low- and middle- income groups.

VII. Comparing the Coefficients from the Two Methods

In Figure 6, we compare the trend in the two sets of estimates in 8th grade math and reading: using the census-derived school income measures and the imputed income measures. The grey lines represent the trend in the overall slope (combining the within- and between-school slopes). The coefficients using imputed income are larger than those based on the census-derived school income measures—likely because they are not attenuated by measurement error and because they include the direct effects of race, mother's education, state, and urbanicity. However, the time trend is very similar: a stable slope during much of the 1990s, with a decline starting in 2005. Although the series for 8th grade reading does not start until 1998, the pattern is very similar, with

the slopes based on imputed income exceeding those based on census-derived income, with both gradually declining over time.

In addition, as shown in Figure 6, we find a similar decline in the within-school coefficient using student-level imputed income as we found with the variance-on-variance regressions: stable during the 1990's and declining after 2005. (The point estimates of the within and between coefficients using imputed income are reported in Appendix Tables G.3 and G.4).

VIII. Reconciling with the Narrowing Racial Gap

Between 1992 and 2015, the Black-White gap in our NAEP data narrowed by one-third of a standard deviation in 4th grade math and by one-fifth of a standard deviation in 8th grade math. Yet the gap in log income by race was stable (see Appendix H). Given the correlation between race and income, the only way that the racial gap and the income-based gap could be moving in different directions would be if the Black-White gap in achievement, conditional on income, were narrowing much faster than the unadjusted Black-White gap. In this section, we ask how much the Black-White gap among students with the same income would have had to change to reconcile a narrowing race gap with the widening achievement-income gap reported by Reardon (2011).

Suppose we had estimates of the bivariate coefficients of achievement on race (β_{1t}) and achievement on income (γ_{1t}) as below:

$$Score_{it} = \beta_{0t} + \beta_{1t}Black_{it} + \varepsilon_{it}$$

$$Score_{it} = \gamma_{0t} + \gamma_{1t}Income_{it} + \vartheta_{it}$$

Now, imagine we were seeking to estimate the conditional effect of race, $\theta_{black,t}$, and income, $\theta_{Income,t}$, in the equation below:

$$Score_{it} = \theta_o + \theta_{Income,t}Income_{it} + \theta_{black,t}Black_{it} + \varphi_{it}$$

To estimate $\theta_{black,t}$ from the bivariate coefficients, β_{1t} and γ_{1t} , we need two auxiliary regressions, one regressing income on a race dummy ($b_{income,black,t}$) and another regressing a race dummy on income ($b_{black,income,t}$). As we show in Appendix H, the Black-White gap conditional on income, $\theta_{black,t}$, can be expressed as a function of the bivariate coefficients, $\hat{\beta}_{1,t}$ and $\hat{\gamma}_{1,t}$, and the auxiliary regression coefficient of income on race ($b_{income,black,t}$) and race on income ($b_{black,income,t}$), both of which we can estimate in the CPS:

$$(6) \hat{\theta}_{black,t} = \hat{\beta}_{1,t} \frac{1}{(1 - b_{black,income,t} b_{income,black,t})} - \hat{\gamma}_{1,t} \frac{b_{income,black,t}}{(1 - b_{black,income,t} b_{income,black,t})}$$

Thus, we use equation (6) to infer the income-adjusted Black-White gap, $\hat{\theta}_{black,t}$, under two different scenarios involving the achievement-income slope ($\hat{\gamma}_{1,t}$): first using our estimated decline and, second, using Reardon's finding of a sharp increase in income-based gaps. In both scenarios, we take as given the unadjusted Black-White gap ($\hat{\beta}_{1,t}$) and the auxiliary regression coefficients ($b_{black,income,t} b_{income,black,t}$). Using our estimates of a declining income-achievement slope for 8th grade math, equation (6) would have implied that the income-adjusted Black-White gap was unchanged, at -.2 standard deviations in both 1992 and 2015, respectively.¹⁶

To simulate the implications of Reardon's 40 percent increase, we increased the achievement-income slope ($\hat{\gamma}_{1,t}$) by 1.6 percent annually between 1992 and 2015 (equivalent to a 40 percent increase over 25 years) and recalculated the income-adjusted Black-White gap. To reconcile Reardon's finding with the observed narrowing of the unadjusted racial gap, the Black-White gap *conditional on income* would have had to change sign from -.22 standard deviations to .19 standard

¹⁶ Recall that the key assumption underlying our method using demographic group mean income and mean achievement required that the direct effect of the groups was unchanging. This would explain why our estimates of the trend were similar using neighborhood incomes as when using the demographic group means.

deviations between 1992 and 2015. The same was true in 4th grade math and reading as well as 8th grade reading (see Appendix H). Thus, in order to be consistent with the observed narrowing of the Black-White gap, a 40 percent increase in the achievement-income slope would have required Black students to be scoring higher than White students with the same income by 2015.

IX. Conclusion

The ideal data for tracking the change in income-based achievement gaps would combine a consistent measure of student achievement with a student-level measure of income over a long period of time. Unfortunately, the federal government has not collected such data. Thus, any effort to shed light on historical changes in income-based achievement gaps will necessarily involve compromises.

In his work, Reardon (2011) prioritized surveys that included a measure of student-level income—even if the surveys were not designed to be comparable, if some income measures were parent-reported, some student reported, and even if the age of the students and the type of assessment varied. Any trend with respect to birth cohort could be driven by the assumptions needed to make the measures comparable. At the least, the confidence intervals around the trend reported by Reardon do not reflect the uncertainty behind those assumptions.

In contrast, we prioritized surveys with a consistent assessment scale, consistent age groups, and consistent methods for inferring income over time. In lieu of student-level income measures, we used school-level measures of income and student-level correlates of income to impute income. Using the two different methods, our findings are quite different from Reardon's—that income-based achievement gaps were narrowing between 1990 and 2015.

Of course, each of our two methods has its own shortcomings. In measuring the mean and variance in incomes by school, we used the population of nearby census tracts and not actual school attendance zones. Thus, an increase in attenuation bias due to measurement error could lead to narrowing income-based gaps in achievement. Nevertheless, the observed correlation between our estimates of school income and administrative data on the percent of students in a school receiving free and reduced-price lunch has not changed. Moreover, both of our corrections for measurement error in the income variable—an analytic adjustment for changing sample size and an instrumental variable estimate using lagged neighborhood income measures—reduced the magnitude of the decline but continued to show that the achievement-income relationship flattened. Thus, measurement error in school-level income is unlikely to account for our finding of narrowing gaps.

The most imprecise component of our estimator is the estimate of the within-school slope using the school-level variance in achievement and the neighborhood variance in income (as described in equation (4).) However, as reported in Figure 6, we find the same narrowing of the within-school coefficient when using student-level imputed income (based on student-level traits varying within school).

Most importantly, our two sets of findings—using school-level aggregates of the mean and variance of income and achievement and imputing income with student-level traits (race/ethnicity, mother’s education, urbanicity, and state)—rely on different assumptions and, yet, they tell the same story. Moreover, our findings are consistent with the narrowing racial achievement gap observed in the NAEP. In contrast, Reardon’s (2011) finding of sharply widening achievement gaps would have required the Black-White gap in achievement conditional on income to have changed sign, with Black students eventually outscoring White students at the same income.

Our results imply that the United States narrowed income-based achievement gaps during a period when higher-income families had greater discretionary income to spend on out-of-school enrichment and despite greater income-based residential segregation. We cannot yet say which policies or social trends were most important in producing such a narrowing in the face of rising income inequality. In some grades and subjects, the implied achievement gains at the 10th percentile of the family income distribution have been remarkable: a full standard deviation in 4th grade math and .7 standard deviations in 8th grade math. In future work, we will be using the estimated shifts in intercepts and slopes of the student level-relationship between achievement and income to sort out the relative effect of state accountability reforms, school finance reforms and other state-level policy variables in explaining the rise in achievement and narrowing of income and race-gaps since 1990.

References:

- Bailey, Martha, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker, “Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program” NBER Working Paper No. 26942, April 2020.
- Bryk, Anthony S., and Stephen W. Raudenbush. *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc, 1992.
- Currie, Janet and Valentina Duque. “Medicaid: What Does It Do, and Can We Do It Better?” *The ANNALS of the American Academy of Political and Social Science* (2019) 686(1), 148–179.
- Currie, Janet and Hannes Schwandt. “Inequality in mortality decreased among the young while increasing for older adults, 1990–2010,” *Science*, Vol. 352, No. 6286, April 2016, pp. 708-712.
- Dee, Thomas S., and Brian Jacob. “The impact of No Child Left Behind on student achievement.” *Journal of Policy Analysis and Management* (2011) 30(3), 418-446.
- Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2020. <https://doi.org/10.18128/D030.V8.0>

Hanushek, Eric A. and Margaret E. Raymond. "Does Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* (2005) 24, pp. 297-327.

Hanushek, Eric A., Paul E. Peterson, Laura Talpey and Ludger Woessmann. "The Unwavering SES Achievement Gap: Trends in U.S. Student Performance." NBER Working Paper No. 25648, March 2019 (Updated March 2020).

Jackson, C. Kirabo, Rucker C. Johnson and Claudia Persico. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *The Quarterly Journal of Economics*, Volume 131, Issue 1, February 2016, Pages 157–218.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. "School Finance Reform and the Distribution of Student Achievement." *American Economic Journal: Applied Economics* 2018, Vol. 10, No. 2, pp. 1-26.

Murnane, Richard J. and Sean F. Reardon. "Long-Term Trends in Private School Enrollments by Family Income" *AERA Open* January-March 2018, Vol. 4, No. 1, pp. 1–24.

National Academies of Sciences, Engineering, and Medicine. *A Roadmap to Reducing Child Poverty* Washington, DC: The National Academies Press, 2019.

National Center for Education Statistics. (2020). Digest of Education Statistics 2018: Chapter 2 - Elementary and Secondary Education; 2020 ASI 4824-2.25005.

Nielsen, Eric. "The Income-Achievement Gap and Adult Outcome Inequality." Unpublished paper, Federal Reserve Board, May 13, 2019.

Owens, Ann, Sean F Reardon, and Christopher Jencks. "Income Segregation Between Schools and School Districts." *American Educational Research Journal* 53, no. 4 (2016): 1159-197.

Raudenbush, Stephen W. and Anthony S. Bryk. Hierarchical linear models: Applications and data analysis methods. Vol. 1. sage, 2002.

Reardon, Sean F. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." in Greg J. Duncan and Richard J. Murnane (eds.) *Whither Opportunity?* (New York: Russell Sage Foundation, 2011).

Reardon, Sean F, and Kendra Bischoff. "Income Inequality and Income Segregation." *AJS; American Journal of Sociology* 116, no. 4 (2011): 1092-153.

Saporito, Salvatore and Deenesh Sohoni "Mapping Educational Inequality: Concentrations of Poverty among Poor and Minority Students in Public Schools" *Social Forces*, Volume 85, Number 3, March 2007.

Watson, Tara. "Inequality and the Measurement of Residential Segregation by Income." *Review of Income and Wealth* 55, no. 3 (2009): 820-44.

Table 1: Correlation Between Census-derived School Characteristics and the Common Core of Data

	1990	2000	2009	2011	2013	2015
Percent of Students Receiving FRPL						
Correlation between CCD and Estimate	N.A.	0.715	0.669	0.655	0.694	0.690
Coefficient on Estimate in Predicting CCD	N.A.	1.042	0.938	0.890	0.926	0.918
<i>N</i>		60,860	74,930	76,400	76,130	75,820
Proportion of Students Black (including Black-Hispanic)						
Correlation between CCD and Estimate	0.853	0.903	0.900	0.899	0.901	0.898
Coefficient on Estimate in Predicting CCD	0.915	0.976	0.977	0.955	0.952	0.957
<i>N</i>	59,470	67,970	76,610	76,770	76,630	76,560
Proportion of Students White						
Correlation between CCD and Estimate	0.828	0.859	0.912	0.914	0.914	0.913
Coefficient on Estimate in Predicting CCD	1.006	1.053	0.982	0.966	0.963	0.967
<i>N</i>	60,796	67,972	76,607	76,773	76,630	76,564
<i>Total # Schools in CCD</i>	68,300	69,790	76,610	76,780	76,630	76,560

Note: 52% schools are missing FRPL information in the CCD in 1990, 13% are missing in 2000, and less than 2% are missing in 2009 and beyond. The CCD and census handle race/ethnicity differently. For both datasets, the Black category includes both Black Hispanics and Black non-Hispanics. In the CCD, the White category always includes both White Hispanics and White non-Hispanics. The census data codes both White Hispanics and White non-Hispanics as "White" in 1990 and 2000. From 2009 on, our census-based estimates include only White non-Hispanics but are still compared to the CCD proportions that include White Hispanics as well.

Table 2: Reconstructing the Relationship Between Student Achievement and Income using School-Level Aggregates

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope	Predicted Score at 10 th Percentile	Predicted Score at 90 th Percentile	Predicted Difference at 90 th vs. 10 th Percentile (6) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
4th Grade Math							
1992	0.867 (0.019)	0.454 (0.041)	0.193	0.533 (0.034)	-0.645	0.480	1.126
2015	0.657 (0.012)	0.179 (0.051)	0.257	0.302 (0.038)	0.353	1.045	0.692
8th Grade Math							
1992	0.892 (0.021)	0.567 (0.039)	0.162	0.620 (0.033)	-0.719	0.589	1.308
2015	0.719 (0.016)	0.349 (0.036)	0.221	0.430 (0.028)	-0.035	0.952	0.986
4th Grade Reading							
1992	0.730 (0.021)	0.528 (0.034)	0.193	0.567 (0.028)	-0.671	0.526	1.197
2015	0.673 (0.012)	0* (0.000)	0.257	0.173 (0.003)	-0.001	0.394	0.396
8th Grade Reading							
1998	0.701 (0.025)	0.429 (0.072)	0.169	0.475 (0.061)	-0.523	0.493	1.016
2015	0.637 (0.015)	0.195 (0.063)	0.221	0.292 (0.049)	-0.209	0.461	0.670

Notes: Composite test scores were standardized with the mean and standard deviation in 1992 (4th grade math and reading, 8th grade math) or 1998 (8th grade reading). For estimates for all years, see Appendix E. The first year of 8th grade reading scores was 1998.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Mean and variance of income were derived from 1990 and 2000 decennial censuses and the American Communities Survey from the Bureau of the Census (ASEC). Income percentiles are based on household income for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

* Point estimate was negative.

Table 3: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates

Year	IV Estimates of the Student-Level Slope				OLS Estimates of the Student-Level Slope			
	Grade 4		Grade 8		Grade 4		Grade 8	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1992	0.697 (0.063)	0.743 (0.065)	0.521 (0.113)		0.533 (0.034)	0.567 (0.028)	0.620 (0.033)	
1998		0.793 (0.077)		0.499 (0.174)		0.602 (0.037)		0.475 (0.061)
2015	0.417 (0.056)	0.189 (0.003)	0.454 (0.055)	0.408 (0.060)	0.302 (0.038)	0.173 (0.003)	0.430 (0.028)	0.292 (0.049)

Notes: For columns (1) through (4), we estimate the between-school coefficient and within-school coefficients separately, using lagged mean income and variance in incomes as instruments in separate specifications. We provide only the student-level coefficient above. The within- and between-school slope coefficients, and the estimates for all years, are reported in Appendix F. Columns (5) through (8) are a reproduction of estimates reported in Table 2. Composite test scores were standardized with the mean and standard deviation in 1992 (4th grade math, 4th grade reading, 8th grade math) or 1998 (8th grade reading). For all years, see Appendix E.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. The instrumental variables--the estimated mean and variance of school income--were estimated using the 1980, 1990 and 2000 decennial censuses.

Table 4: The Relationship Between Achievement and Alternate Ways of Imputing Income

Year	Student Traits Used			
	Race/ Ethnicity	Mother's Education	Urbanicity and State	Race*Moth ed, Urbanicity, State
8th Grade Math				
1992	1.435 (0.024)	0.749 (0.015)	0.805 (0.044)	0.938 (0.015)
2015	1.186 (0.018)	0.677 (0.011)	0.517 (0.036)	0.804 (0.009)
8th Grade Reading				
1998	1.048 (0.032)	0.606 (0.016)	0.472 (0.046)	0.753 (0.017)
2015	1.081 (0.018)	0.597 (0.013)	0.478 (0.032)	0.704 (0.011)

Notes: Each column reports the slope of a student-level regression of 8th grade math scores on imputed income, with the imputations based on the variables indicated at the top of each column. Composite test scores were standardized with the mean and standard deviation in 1992 (math) or 1998 (reading). Student traits only available for 8th grade students. For estimates for all years, see Appendix Tables G.1 and G.2.

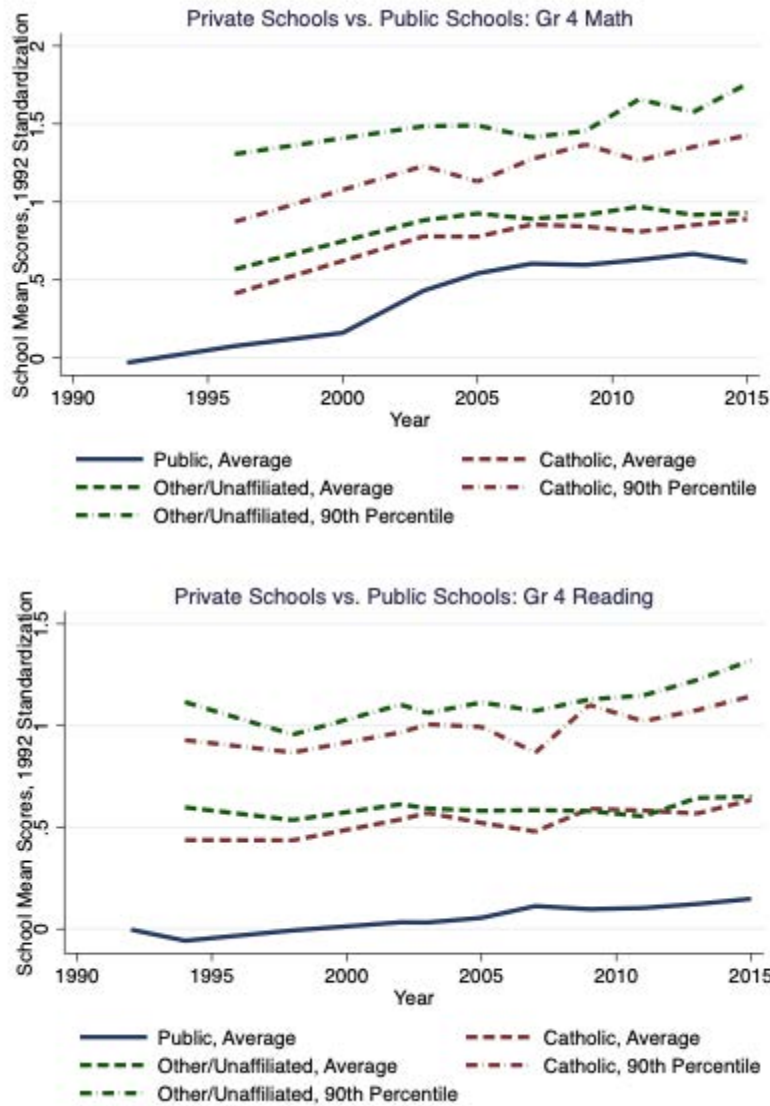
Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org. Because the main NAEP assessments are collected between January and March, we use household income from the prior calendar year.

Appendix A: Could Private School Achievement Have Led to a Widening Gap?

Our main analyses are limited to public schools. Yet, as reported by Murnane and Reardon (2018), children in families at the 90th income percentile are roughly 4 times as likely to enroll in a private school than students at the 10th percentile of family income (roughly 18 percent versus 4 percent during the time period we are studying). So, the question is, might the rise in the gap in achievement by income that Reardon (2011) reported be driven by a sharp widening of the gap in achievement between public schools and private schools?

There is little reason to believe that our conclusions would be any different if we had been able to include student-level data on income and achievement from private schools. In Figure A.1, we report the trend in achievement for the average public school and the average private school for grade 4 reading and math. We also report the achievement of the 90th percentile private school, recognizing that high-income families are unlikely to attend the average private school. For each grade-subject combination, the public-private gap in mean achievement has been narrowing over time. In addition, the gap between the average public school and the 90th percentile private school has also been narrowing. Figure A.2 reports similar results for 8th grade students. Again, there is no evidence of widening.

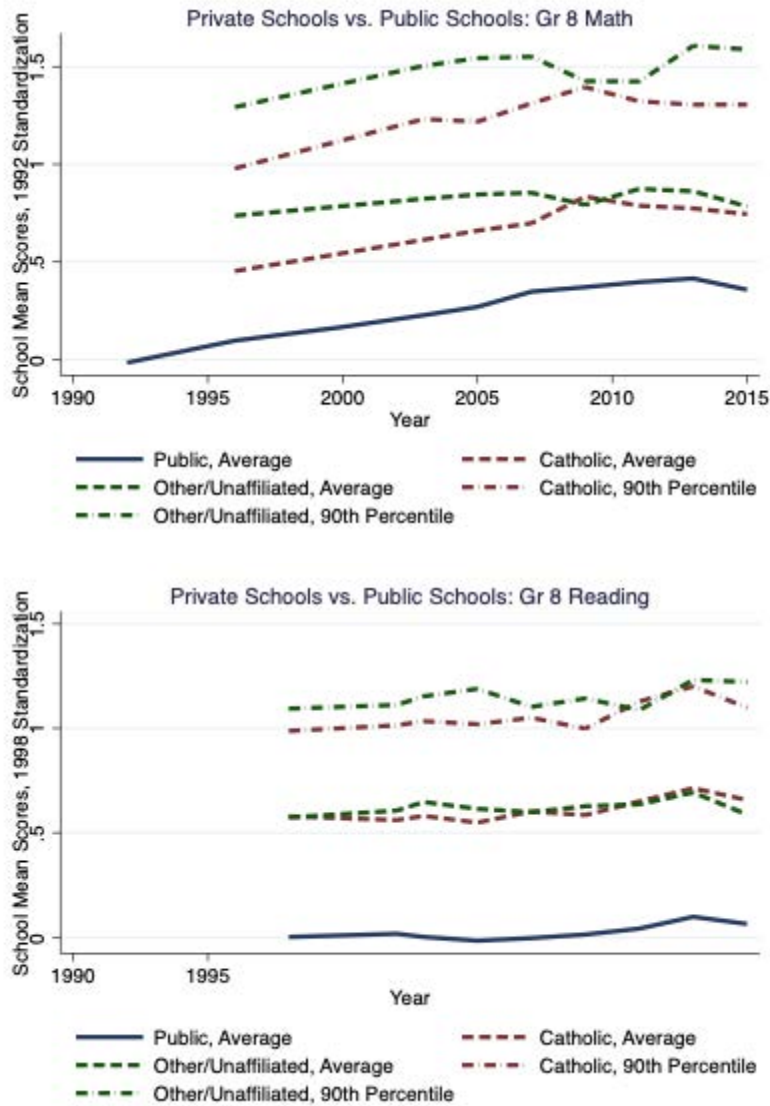
Appendix Figure A.1: Trend in Achievement in Private Schools (4th Grade Math and Reading)



Notes: For each school type, school-level weighted averages are calculated using NAEP’s student-level weights. The average and 90th percentile of these school-level estimates are plotted for each year where school type is available.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015.

Appendix Figure A.2: Trend in Achievement in Private Schools (8th Grade Math and Reading)



Notes: For each school type, school-level weighted averages are calculated using NAEP's student-level weights. The average and 90th percentile of these school-level estimates are plotted for each year where school type is available.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015.

Appendix B: Reconciling with Research on Achievement Gaps by SES

In a recent paper, Hanushek, Peterson, Talpey and Woessmann (2020) use the NAEP (along with the TIMSS and PISA exams) to study trends in SES gaps over time. In this section, we replicate their methods for calculating the SES index to reconcile their findings regarding SES-based gaps with our findings regarding income-based gaps in the NAEP. After doing a principal component analysis (separately each year), we follow the authors in using the factor loadings from the first factor to create an SES index using mother's education, father's education, and student reports regarding a list of up to 10 household possessions.

We use the resulting SES index in two ways: first, we use it to validate our approach to estimating the within-school slope using school-level variance-on-variance regression to what we find with student-level data. (Unfortunately, we cannot do the same test with income since we don't have student-level data on income. Nevertheless, if there were factors biasing our estimate of the within-school achievement-income slope over time, then we might expect them to be biasing our within-school achievement-SES slope as well). To test the validity of our approach, we first regress school-level variance in achievement on the variance in SES in the school. Second, we estimate the within-school coefficient of achievement on SES with student-level data (including school fixed effects). (The results by year for 8th grade math and reading are reported in Appendix Table B.1). In Appendix Figure B.1, we report the trend in the within-school slope implied by the school-level variance-on-variance regression and by the student-level regressions. The within-school slope from the school-level regression is biased upward (because it contains variation in the within-school slope between schools as discussed in the methods section). However, the trend over time is largely parallel to the trend estimated with student-level data.

Second, we use the SES index to reconcile our finding of narrowing gaps with respect to income with Hanushek et al.'s (2020) finding of stable or widening gaps with respect to SES. The

first column of Appendix Tables B.2 and B.3 reports the student-level coefficient of achievement on SES that we estimate in the NAEP for 8th grade math and reading respectively. As reported in column (1), the gaps in achievement by SES were fairly stable between 1992 and 2000 but increased slightly between 2002 and 2015. The second column reports the coefficient of imputed log income on SES. The coefficient is rising steadily throughout the period, rising from .282 to .376 in math (a 33 percent increase) as we show in Appendix Table B.2. In other words, a unit change in the SES index represented a larger change in income over time. The third column is the ratio of the first two. Analogous to the indirect least squares when using SES as the instrument, the ratio is falling over time in math and reading (although the decline starts later in reading). In other words, the widening in the achievement-SES gap is consistent with a narrowing of income-based achievement gaps, primarily because a unit change in SES was associated with a larger difference in income.

Appendix Table B.1: Within-School Estimates of the Relationship Between Achievement and SES

Year	Grade 8 Math		Grade 8 Reading	
	Regressing Var Score on Var SES Index	Regressing Score on SES with School FEs	Regressing Var Score on Var SES Index	Regressing Score on SES with School FEs
1992	0.347 (0.021)	0.225 (0.006)		
1994				
1996	0.286 (0.029)	0.206 (0.006)		
1998			0.350 (0.026)	0.209 (0.006)
2000	0.289 (0.035)	0.217 (0.008)		
2002			0.301 (0.027)	0.251 (0.006)
2003	0.315 (0.020)	0.269 (0.006)	0.389 (0.016)	0.258 (0.005)
2005	0.331 (0.018)	0.276 (0.004)	0.372 (0.017)	0.271 (0.004)
2007	0.307 (0.021)	0.272 (0.005)	0.404 (0.016)	0.264 (0.005)
2009	0.299 (0.021)	0.267 (0.006)	0.380 (0.016)	0.260 (0.006)
2011	0.262 (0.024)	0.283 (0.006)	0.322 (0.019)	0.272 (0.006)
2013	0.363 (0.020)	0.321 (0.005)	0.448 (0.016)	0.314 (0.005)
2015	0.373 (0.020)	0.320 (0.006)	0.468 (0.017)	0.309 (0.006)

Notes: Columns (1) and (3) are estimates of the within-school slope based on square root of the slope coefficient from the regression of school-level variance in achievement on variance in the SES index. Columns (2) and (4) are estimates of the within-school slope based on the regression of student-level achievement on the SES index. Following Hanushek et al. (2020), the SES index is based on the factor loadings of the first factor from a principal component analysis using parental education and a list of up to 10 reported household possessions in each year and then standardized using the mean and standard deviation in each year. Composite test scores were standardized with the mean and standard deviation in 1992 (math) or 1998 (reading).

Source: Student test score data and data used to create the SES index are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015 and corresponding Student Survey Questionnaires. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table B.2: The Relationship Between Achievement, Income, and SES (8th Grade Math)

Year	Regressing Score on SES Index: $\theta_{Score, SES}$	Regressing Ln Income on SES Index: $\gamma_{Income, SES}$	Implied Coefficient of Score on Income: $\theta_{Score, SES} / \gamma_{Income, SES}$	Student-Level Coefficient: $\beta_{Score, Income}$
1992	0.339 (0.006)	0.282 (0.002)	1.203 (0.019)	0.938 (0.015)
1994				
1996	0.335 (0.007)	0.297 (0.002)	1.127 (0.021)	0.938 (0.015)
1998				
2000	0.331 (0.010)	0.289 (0.003)	1.144 (0.032)	0.932 (0.026)
2002				
2003	0.376 (0.005)	0.310 (0.002)	1.212 (0.016)	0.870 (0.010)
2005	0.386 (0.004)	0.314 (0.001)	1.228 (0.013)	0.916 (0.010)
2007	0.384 (0.005)	0.337 (0.002)	1.138 (0.015)	0.827 (0.009)
2009	0.386 (0.006)	0.357 (0.002)	1.080 (0.018)	0.797 (0.010)
2011	0.390 (0.008)	0.369 (0.002)	1.058 (0.021)	0.766 (0.010)
2013	0.425 (0.005)	0.378 (0.002)	1.124 (0.013)	0.776 (0.008)
2015	0.442 (0.005)	0.376 (0.002)	1.176 (0.015)	0.804 (0.009)

Notes: Composite test scores were standardized with the mean and standard deviation in 1992. Following Hanushek et al. (2020), the SES index is based on the factor loadings of the first factor from a principal component analysis using parental education and a list of up to 10 reported household possessions in each year and then standardized using the mean and standard deviation in each year.

Source: Student test score data and data used to create the SES index are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015 and corresponding Student Survey Questionnaires. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

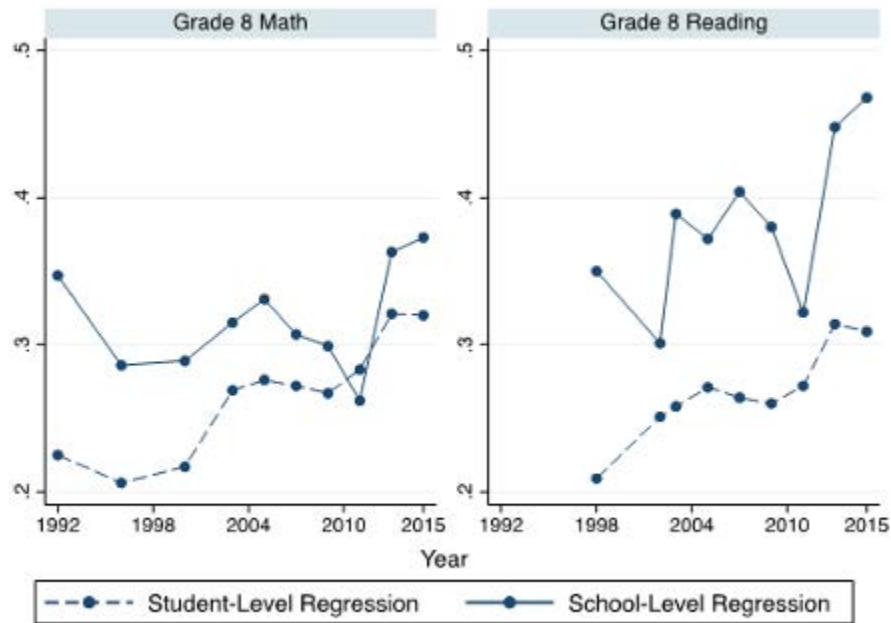
Appendix Table B.3: The Relationship Between Achievement, Income, and SES (8th Grade Reading)

Year	Regressing Score on SES Index: $\theta_{Score, SES}$	Regressing Ln Income on SES Index: $\gamma_{Income, SES}$	Implied Coefficient of Score on Income: $\theta_{Score, SES} / \gamma_{Income, SES}$	Student-Level Coefficient: $\beta_{Score, Income}$
1992				
1994				
1996				
1998	0.292 (0.006)	0.304 (0.003)	0.958 (0.020)	0.753 (0.017)
2000				
2002	0.325 (0.006)	0.231 (0.002)	1.406 (0.027)	0.728 (0.014)
2003	0.353 (0.006)	0.310 (0.002)	1.139 (0.018)	0.792 (0.011)
2005	0.360 (0.004)	0.309 (0.001)	1.165 (0.012)	0.800 (0.008)
2007	0.356 (0.005)	0.330 (0.002)	1.078 (0.015)	0.742 (0.009)
2009	0.354 (0.006)	0.352 (0.002)	1.005 (0.016)	0.722 (0.011)
2011	0.362 (0.005)	0.369 (0.002)	0.981 (0.014)	0.694 (0.010)
2013	0.408 (0.005)	0.380 (0.002)	1.076 (0.014)	0.705 (0.009)
2015	0.409 (0.006)	0.375 (0.002)	1.089 (0.017)	0.704 (0.011)

Notes: Composite test scores were standardized with the mean and standard deviation in 1998. Following Hanushek et al. (2020), the SES index is based on the factor loadings of the first factor from a principal component analysis using parental education and a list of up to 10 reported household possessions in each year and then standardized using the mean and standard deviation in each year.

Source: Student test score data and data used to create the SES index are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015 and corresponding Student Survey Questionnaires. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Figure B.1: Comparing Trends in Within-School Estimates of the Relationship Between Achievement and SES



Notes: The line graphs portray the estimated within-school coefficient of NAEP standardized scale score on socioeconomic status (SES). The school-level estimates are based on the regression of school-level variance in achievement on variance in SES. Following Hanushek et al. (2020), the SES index is based on the factor loadings of the first factor from a principal component analysis using parental education and a list of up to 10 reported household possessions in each year and then standardized using the mean and standard deviation in each year. The student-level estimates are based on student-level data including school fixed effects. Composite test scores were standardized with the mean and standard deviation in 1992 (math) or 1998 (reading).

Source: Student test score data and data used to create the SES index are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015 and corresponding Student Survey Questionnaires.

Appendix C: Analytic Adjustment for Attenuation Bias Due to Measurement Error

In this appendix we describe how we calculate an analytic correction for the attenuation bias due to measurement error in the school-level data on average log income (\bar{Y}_j) and variance of log income ($\hat{\sigma}_{Y_j}^2$). Recall that we obtain estimates of $\beta_{between}$ and β_{within}^2 from OLS estimates of equations (3) and (4):

$$(3) \quad \overline{Score}_j = \beta_0 + \beta_{between}\bar{Y}_j + \varepsilon_j$$

$$(4) \quad \sigma_v^2 + \beta_{within}^2 \hat{\sigma}_{Y_j}^2 + \vartheta_j$$

The independent variables, \bar{Y}_j and $\hat{\sigma}_{Y_j}^2$, are derived from a sample of households in each census tract, and sampling error in these estimates will result in attenuation bias. Given that the 1990 census sent long forms to a higher percentage of households than either the 2000 census or the ACS, we explore the implications of changing sample size for changes in attenuation bias. Let $\bar{Y}_j = \mu_j + u_j$ and $\hat{\sigma}_{Y_j}^2 = \sigma_{Y_j}^2 + e_j$, where u_j and e_j represent the estimation error in each of the estimates. Then the plim of OLS estimates of equations (3) and (4) will be:

$$(A1) \quad \text{plim } \hat{\beta}_{between} = \delta_{\bar{Y}} \beta_{between}, \text{ where } \delta_{\bar{Y}} = 1 - \frac{\text{Var}(u_j)}{\text{Var}(\bar{Y}_j)} \text{ is the reliability of } \bar{Y}_j$$

$$(A2) \quad \text{plim } \hat{\beta}_{within}^2 = \delta_{\hat{\sigma}_{Y_j}^2} \beta_{within}^2, \text{ where } \delta_{\hat{\sigma}_{Y_j}^2} = 1 - \frac{\text{Var}(e_j)}{\text{Var}(\hat{\sigma}_{Y_j}^2)} \text{ is the reliability of } \hat{\sigma}_{Y_j}^2$$

Therefore, to dis-attenuate the slope coefficients in equations (3) and (4) we divide the OLS

estimate by the corresponding reliability of the regressor in each equation ($\frac{\hat{\beta}_{between}}{\delta_{\bar{Y}}}, \frac{\hat{\beta}_{within}^2}{\delta_{\hat{\sigma}_{Y_j}^2}}$).

We derive the reliability of our estimator of average log income ($\delta_{\bar{Y}} = 1 - \frac{Var(u_j)}{Var(\bar{Y}_j)}$) and our

estimator of the variance of log income ($\delta_{\hat{\sigma}_{Y_j}^2} = 1 - \frac{Var(e_j)}{Var(\hat{\sigma}_{Y_j}^2)}$) as follows:

1. We calculate the total variance of \bar{Y}_j and $\hat{\sigma}_{Y_j}^2$ directly from the school-level data used to estimate equations (3) and (4).
2. We calculate the variance of the estimation error in \bar{Y}_j using the usual formula for variance of a sample average, where $Var(u_j) = \frac{\sigma_{Y,within}^2}{n}$, and $\sigma_{Y,within}^2 = (1 - ICC)\sigma_Y^2$. The latter is the average within-school variance in log income and n is the number of households used to estimate school-level income on average, while *ICC* is the intra-class correlation in income by school.
3. We calculate the variance of the estimation error in $\hat{\sigma}_{Y_j}^2$ using the formula for the variance of the sample variance for a normally distributed variable, where $Var(e_j) = \frac{2(\sigma_{Y,within}^2)^2}{n}$, and $\sigma_{Y,within}^2$ is the average within-school variance in log income and n is the number of households used to estimate school-level income on average.

This approach yields a measure of the reliability of the income measures for years in which we have data (the 1990 census, the 2000 census, and the 2009-2015 American Community Survey). For the remaining years, we linearly interpolate the average log income and variance of log income, so we are using weighted averages of the two nearest census-year estimates. To calculate reliability of our interpolated estimates, we calculate total variance directly as in step

(1) above. Because estimation error is independent across census years, the variance of the estimation error for any weighted average is simply

$var(wX_1 + (1 - w)X_2) = w^2 var(X_1) + (1 - w)^2 var(X_2)$, where X_1 and X_2 are the two census-year estimates, w is the weight placed on the first census-year estimate, and $var(X_1)$ and $var(X_2)$ are the variances of the estimation error of the census-year estimates calculated as described in steps (2) and (3) above.

Since our estimates of school-level means and variances of log income are derived from binned data (using a regression as described in the text), rather than derived from continuous income data, it is possible that these formulas may overstate the reliability of the census estimates. Therefore, as a check, we also calculated the variance of the estimation error (steps (2) and (3) above) using an alternative method: we calculate the variance of the estimation error in \bar{Y}_j using the estimated standard error on this estimate from the regressions, where

$Var(u_j) = \overline{SE_{\bar{Y}}^2}$ is the sampling variance (the standard error squared) of our estimate of each school's average log income averaged across all schools in the sample. Similarly, we calculate the variance of the estimation error in $\hat{\sigma}_{Y_j}^2$ using the estimated standard error on this estimate from the regressions, where $Var(e_j) = \overline{SE_{\hat{\sigma}_{Y_j}^2}^2}$ is the sampling variance (the standard error squared) of our estimate of each school's variance of log income averaged across all schools in the sample. This method yielded results that were very similar to the analytic correction discussed above.

Appendix D: Data and Sample Sizes

Appendix Table D.1 reports the R^2 and sample size for a separate regression predicting income in the Annual Social and Economic Supplement to the Current Population Survey in each year as described by equation (5).

We report the number of schools and students by grade/subject and by year in Appendix Table D.2. For example, in 2015, we included scores from 104,400 students attending 5,460 schools in 4th grade reading.

Estimating the Mean and Variance in Income for Schools. The block group data report the number of related children attending public schools in grades 1-4 and in grades 5-8 in each block group. Using these, we identify the K nearest block groups with a sufficient number of children to fill up the school's official enrollment. In other words, if N_j is the number of students enrolled in school j according to the CCD, then we associate the nearest K block groups that satisfy the inequality:

$$\sum_{k=1}^{K-1} N_k < N_j \leq \sum_{k=1}^K N_k$$

where block groups k are ordered in increasing distance from the school location.

In the census, detailed data on household income are only available at the census tract level not the block-group level. As a result, we assume that the distribution of income in each block group matches the distribution in its associated census tract. We then weight the tract-level data by the counts of students in each of the K block groups associated with the school. Thus, the number of students in school j who come from a household with income in the i^{th} reported bin is given by:

$$N_{ji} = \sum_{k=1}^K \frac{p_{ki} n_k}{n_j}$$

where p_{ki} is the proportion of students in the corresponding census tract in income bin i . These counts can be thought of as a coarsened distribution of the household income for students in each school.

To calculate the mean and variance of log income for each school, we use a method outlined by vonHippel, Scarpino, and Drown (2016), fitting each school's income bins to a log-normal distribution. If the log of income, $\ln(Y)$, is distributed normally, and C_l is the upper range of the l th income bin (expressed in log 2016 dollars), then the proportion of the school enrollment with income below can be expressed as follows:

$$P_{C_l} = \Phi \left(\frac{\ln(C_l) - \mu_{Y_j}}{\sigma_{Y_j}} \right)$$

where μ_{Y_j} and $\sigma_{Y_j}^2$ represent the mean and variance of log income in school j .

Accordingly, using for the top limit in each income bin, C_l (where $l=1, \dots, 15$ corresponds to 16 income bins) and the proportion of households, P_{C_l} , in the school with incomes below C_l , we estimate the following regression for each school:

$$\ln(C_l) = \mu_{Y_j} + \sigma_{Y_j} \text{invnorm}(P_{C_{jl}}) + \vartheta_{jl}$$

where j subscripts the school and l subscripts one of 15 income bins.

Appendix Table D.1: R2 for First-Stage Imputations of Household Income Using CPS

Year	R ²	CPS Sample Size
1992	0.267	32,230
1993	0.271	32,203
1994	0.287	31,879
1995	0.258	31,910
1996	0.250	28,287
1997	0.256	28,486
1998	0.259	28,513
1999	0.260	28,704
2000	0.255	28,695
2001	0.235	53,812
2002	0.246	53,339
2003	0.249	52,741
2004	0.243	51,707
2005	0.235	50,644
2006	0.247	49,797
2007	0.260	48,738
2008	0.256	48,067
2009	0.252	47,938
2010	0.261	47,689
2011	0.277	45,764
2012	0.261	44,794
2013	0.268	45,118
2014	0.243	44,154
2015	0.240	43,660

Notes: In columns (2) and (3), each row presents the R2 and sample size for a separate regression predicting income in the CPS data in that year. The increase in sample size in 2001 is due to a sample expansion to the ASEC.

Source: The Annual Social and Economic Supplement of the Current Population Survey (ASEC) from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table D.2: NAEP Administration Schedule and State Sample Size

Year	Grade 4 Reading		Grade 8 Reading		Grade 4 Math		Grade 8 Math	
	Schools	Students	Schools	Students	Schools	Students	Schools	Students
1992	4,020	94,720			4,110	97,680	3,250	92,910
1994	3,660	94,350						
1996					3,770	92,000	3,210	82,330
1998	3,220	77,900	2,840	73,020				
2000					3,280	72,060	2,830	71,020
2002	4,080	109,940	3,860	100,120				
2003	5,200	137,840	4,630	121,970	5,630	151,740	4,440	119,640
2005	6,520	122,370	5,340	126,600	7,020	135,010	5,150	124,180
2007	5,460	141,360	5,350	129,500	5,870	155,970	5,110	119,240
2009	6,690	132,810	5,380	130,430	7,160	133,490	5,130	125,690
2011	5,750	156,860	5,540	131,440	6,220	163,310	5,270	131,390
2013	5,670	142,290	5,200	140,470	6,140	148,580	4,990	134,390
2015	5,460	104,400	4,780	111,880	5,890	111,310	4,580	108,170
<i>States in our sample</i>	42		46		45		45	

Notes: School counts include District of Columbia, exclude Department of Defense Schools, and exclude Bureau of Indian Affairs schools. States were included if they opted out of no more than one available NAEP administration during the period between 1992 and 2002. All counts were rounded to the nearest 10, per IES reporting requirements.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015.

Appendix E: Reconstructing the Relationship Between Student Achievement and Income Using School-Level Aggregates

Appendix Table E.1: Reconstructing the Relationship Between Student Achievement and Income Using School-Level Aggregates (4th Grade Math)

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope	Predicted Score at 10 th Percentile	Predicted Score at 90 th Percentile	Predicted Difference at 90 th vs. 10th Percentile (6) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992	0.867 (0.019)	0.454 (0.041)	0.193	0.533 (0.034)	-0.645	0.480	1.126
1994							
1996	0.904 (0.021)	0.410 (0.047)	0.195	0.506 (0.038)	-0.477	0.594	1.071
1998							
2000	0.888 (0.022)	0.297 (0.062)	0.215	0.424 (0.049)	-0.247	0.640	0.886
2002							
2003	0.805 (0.013)	0.313 (0.034)	0.220	0.421 (0.027)	0.050	0.933	0.883
2005	0.729 (0.012)	0.272 (0.041)	0.227	0.376 (0.032)	0.183	0.991	0.808
2007	0.723 (0.011)	0.205 (0.054)	0.238	0.328 (0.041)	0.296	1.007	0.710
2009	0.725 (0.011)	0.227 (0.040)	0.250	0.351 (0.030)	0.270	1.032	0.762
2011	0.678 (0.010)	0.211 (0.039)	0.252	0.329 (0.029)	0.288	1.032	0.743
2013	0.669 (0.010)	0.214 (0.039)	0.254	0.330 (0.029)	0.340	1.093	0.753
2015	0.657 (0.012)	0.179 (0.051)	0.257	0.302 (0.038)	0.353	1.045	0.692

Notes: Composite test scores were standardized with the mean and standard deviation in 1992.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Mean and variance of income were derived from 1990 and 2000 decennial censuses and the American Communities Survey from the Bureau of the Census (ASEC). Income percentiles are based on household income for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table E.2: Reconstructing the Relationship Between Student Achievement and Income Using School-Level Aggregates (4th Grade Reading)

	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope	Predicted Score at 10 th Percentile	Predicted Score at 90 th Percentile	Predicted Difference at 90 th vs. 10 th Percentile (6) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992	0.730 (0.021)	0.528 (0.034)	0.193	0.567 (0.028)	-0.671	0.526	1.197
1994	0.837 (0.023)	0.605 (0.047)	0.191	0.649 (0.039)	-0.887	0.548	1.436
1996							
1998	0.836 (0.022)	0.542 (0.045)	0.204	0.602 (0.037)	-0.704	0.582	1.286
2000							
2002	0.860 (0.017)	0.325 (0.053)	0.217	0.441 (0.042)	-0.383	0.542	0.926
2003	0.830 (0.015)	0.358 (0.047)	0.220	0.462 (0.037)	-0.424	0.544	0.968
2005	0.753 (0.013)	0.240 (0.062)	0.227	0.357 (0.048)	-0.308	0.459	0.766
2007	0.741 (0.013)	0.166 (0.089)	0.238	0.302 (0.067)	-0.185	0.470	0.654
2009	0.721 (0.012)	0.152 (0.085)	0.250	0.295 (0.064)	-0.184	0.455	0.639
2011	0.692 (0.011)	0.027 (0.444)	0.252	0.195 (0.331)	-0.087	0.355	0.441
2013	0.696 (0.011)	0* (0.000)	0.254	0.177 (0.003)	-0.033	0.371	0.404
2015	0.673 (0.012)	0* (0.000)	0.257	0.173 (0.003)	-0.001	0.394	0.396

Notes: Composite test scores were standardized with the mean and standard deviation in 1992.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Mean and variance of income were derived from 1990 and 2000 decennial censuses and the American Communities Survey from the Bureau of the Census (ASEC). Income percentiles are based on household income for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

* Point estimate was negative.

Appendix Table E.3: Reconstructing the Relationship Between Student Achievement and Income Using School-Level Aggregates (8th Grade Math)

	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope	Predicted Score at 10 th Percentile	Predicted Score at 90 th Percentile	Predicted Difference at 90 th vs. 10 th Percentile (6) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992	0.892 (0.021)	0.567 (0.039)	0.162	0.620 (0.033)	-0.719	0.589	1.308
1994							
1996	0.922 (0.028)	0.560 (0.046)	0.160	0.618 (0.039)	-0.590	0.719	1.309
1998							
2000	0.909 (0.028)	0.536 (0.061)	0.177	0.602 (0.051)	-0.429	0.828	1.258
2002							
2003	0.897 (0.018)	0.464 (0.039)	0.180	0.542 (0.032)	-0.282	0.854	1.136
2005	0.840 (0.018)	0.507 (0.033)	0.186	0.569 (0.026)	-0.291	0.932	1.222
2007	0.790 (0.017)	0.393 (0.044)	0.199	0.472 (0.035)	-0.102	0.919	1.021
2009	0.803 (0.017)	0.430 (0.033)	0.211	0.509 (0.026)	-0.099	1.004	1.102
2011	0.714 (0.017)	0.330 (0.040)	0.215	0.412 (0.031)	-0.025	0.908	0.932
2013	0.716 (0.016)	0.349 (0.034)	0.217	0.429 (0.026)	-0.004	0.976	0.980
2015	0.719 (0.016)	0.349 (0.036)	0.221	0.430 (0.028)	-0.035	0.952	0.986

Notes: Composite test scores were standardized with the mean and standard deviation in 1992.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Mean and variance of income were derived from 1990 and 2000 decennial censuses and the American Communities Survey from the Bureau of the Census (ASEC). Income percentiles are based on household income for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table E.4: Reconstructing the Relationship Between Student Achievement and Income Using School-Level Aggregates (8th Grade Reading)

Year	Between-School Slope	Within-School Slope	ICCY	Implied Student-Level Slope	Predicted Score at 10 th Percentile	Predicted Score at 90 th Percentile	Predicted Difference at 90 th vs. 10 th Percentile (6) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992							
1994							
1996							
1998	0.701 (0.025)	0.429 (0.072)	0.169	0.475 (0.061)	-0.523	0.493	1.016
2000							
2002	0.727 (0.019)	0.474 (0.042)	0.178	0.519 (0.035)	-0.465	0.624	1.089
2003	0.778 (0.018)	0.366 (0.054)	0.180	0.440 (0.044)	-0.427	0.496	0.923
2005	0.701 (0.016)	0.414 (0.042)	0.186	0.467 (0.034)	-0.498	0.506	1.004
2007	0.691 (0.016)	0.288 (0.058)	0.199	0.368 (0.046)	-0.362	0.433	0.796
2009	0.676 (0.015)	0.284 (0.051)	0.211	0.367 (0.040)	-0.328	0.467	0.795
2011	0.641 (0.016)	0.256 (0.050)	0.215	0.339 (0.039)	-0.304	0.462	0.767
2013	0.657 (0.014)	0.254 (0.045)	0.217	0.342 (0.035)	-0.229	0.552	0.781
2015	0.637 (0.015)	0.195 (0.063)	0.221	0.292 (0.049)	-0.209	0.461	0.670

Notes: Composite test scores were standardized with the mean and standard deviation in 1998.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Mean and variance of income were derived from 1990 and 2000 decennial censuses and the American Communities Survey from the Bureau of the Census (ASEC). Income percentiles are based on household income for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table E.5: Log Household Income Percentiles from March CPS

Year	90th Percentile	10th Percentile	Difference
1992	9.750	11.860	2.111
1993	9.759	11.880	2.120
1994	9.697	11.908	2.211
1995	9.756	11.922	2.166
1996	9.819	11.937	2.118
1997	9.826	11.959	2.133
1998	9.859	11.997	2.138
1999	9.903	12.014	2.110
2000	9.951	12.041	2.090
2001	10.000	12.065	2.065
2002	9.971	12.069	2.097
2003	9.948	12.045	2.096
2004	9.922	12.065	2.143
2005	9.908	12.057	2.149
2006	9.892	12.075	2.183
2007	9.924	12.087	2.163
2008	9.909	12.059	2.150
2009	9.913	12.080	2.168
2010	9.832	12.070	2.239
2011	9.776	12.037	2.262
2012	9.783	12.047	2.264
2013	9.772	12.057	2.285
2014	9.796	12.093	2.298
2015	9.833	12.125	2.292

Notes: Income percentiles are reported in ln 2016 dollars.

Source: Income percentiles are based on household income for households with related children aged 5 to 18 in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix F: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates

Appendix Table F.1: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates (4th Grade Math)

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope
	(1)	(2)	(3)	(4)
1992	0.857 (0.035)	0.659 (0.078)	0.193	0.697 (0.063)
1994				
1996	0.855 (0.038)	0.413 (0.135)	0.195	0.499 (0.109)
1998				
2000	0.877 (0.025)	0.321 (0.111)	0.215	0.441 (0.088)
2002				
2003	0.803 (0.015)	0.347 (0.058)	0.220	0.447 (0.045)
2005	0.731 (0.013)	0.372 (0.064)	0.227	0.453 (0.050)
2007	0.754 (0.014)	0.504 (0.053)	0.238	0.564 (0.041)
2009	0.789 (0.012)	0.398 (0.049)	0.250	0.496 (0.037)
2011	0.744 (0.012)	0.412 (0.048)	0.252	0.496 (0.036)
2013	0.739 (0.012)	0.321 (0.062)	0.254	0.428 (0.047)
2015	0.728 (0.013)	0.310 (0.075)	0.257	0.417 (0.056)

Notes: Estimates of the mean and variance in neighborhood income during the prior decennial census are used to instrument for each year's values. Composite test scores were standardized with the mean and standard deviation in 1992.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. The instrumental variables—the estimated mean and variance of school income—were estimated using the 1980, 1990 and 2000 decennial censuses.

Appendix Table F.2: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates (4th Grade Reading)

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope
	(1)	(2)	(3)	(4)
1992	0.711 (0.035)	0.751 (0.080)	0.193	0.743 (0.065)
1994	0.745 (0.042)	0.897 (0.105)	0.191	0.868 (0.086)
1996				
1998	0.690 (0.045)	0.820 (0.095)	0.204	0.793 (0.077)
2000				
2002	0.862 (0.020)	0.439 (0.067)	0.217	0.531 (0.053)
2003	0.829 (0.018)	0.470 (0.063)	0.220	0.549 (0.050)
2005	0.755 (0.015)	0.336 (0.086)	0.227	0.431 (0.067)
2007	0.753 (0.015)	0.600 (0.054)	0.238	0.637 (0.042)
2009	0.777 (0.013)	0.274 (0.101)	0.250	0.400 (0.076)
2011	0.756 (0.013)	0.379 (0.081)	0.252	0.474 (0.061)
2013	0.776 (0.013)	0.337 (0.087)	0.254	0.449 (0.065)
2015	0.734 (0.014)	0* (0.000)	0.257	0.189 (0.003)

Notes: Estimates of the mean and variance in neighborhood income during the prior decennial census are used to instrument for each year's values. Composite test scores were standardized with the mean and standard deviation in 1992.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. The instrumental variables—the estimated mean and variance of school income—were estimated using the 1980, 1990 and 2000 decennial censuses.

* Point estimate was negative.

Appendix Table F.3: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates (8th Grade Math)

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope
	(1)	(2)	(3)	(4)
1992	0.816 (0.037)	0.464 (0.134)	0.162	0.521 (0.113)
1994				
1996	0.782 (0.045)	0.471 (0.142)	0.160	0.521 (0.119)
1998				
2000	0.889 (0.032)	0.622 (0.084)	0.177	0.669 (0.070)
2002				
2003	0.867 (0.020)	0.478 (0.060)	0.180	0.548 (0.050)
2005	0.818 (0.020)	0.574 (0.052)	0.186	0.619 (0.042)
2007	0.777 (0.019)	0.433 (0.084)	0.199	0.501 (0.067)
2009	0.863 (0.020)	0.553 (0.049)	0.211	0.619 (0.039)
2011	0.762 (0.020)	0.396 (0.069)	0.215	0.475 (0.054)
2013	0.763 (0.018)	0.453 (0.051)	0.217	0.520 (0.040)
2015	0.766 (0.019)	0.366 (0.070)	0.221	0.454 (0.055)

Notes: Estimates of the mean and variance in neighborhood income during the prior decennial census are used to instrument for each year's values. Composite test scores were standardized with the mean and standard deviation in 1992.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. The instrumental variables—the estimated mean and variance of school income—were estimated using the 1980, 1990 and 2000 decennial censuses.

Appendix Table F.4: Instrumenting the Relationship Between Student Achievement and Income using School-Level Aggregates (8th Grade Reading)

Year	Between-School Slope	Within-School Slope	ICC _Y	Implied Student-Level Slope
	(1)	(2)	(3)	(4)
1992				
1994				
1996				
1998	0.574 (0.041)	0.484 (0.209)	0.169	0.499 (0.174)
2000				
2002	0.685 (0.021)	0.552 (0.053)	0.178	0.575 (0.044)
2003	0.744 (0.019)	0.542 (0.058)	0.180	0.579 (0.048)
2005	0.668 (0.019)	0.618 (0.049)	0.186	0.627 (0.040)
2007	0.677 (0.018)	0.501 (0.073)	0.199	0.536 (0.059)
2009	0.722 (0.017)	0.555 (0.048)	0.211	0.590 (0.038)
2011	0.686 (0.019)	0.431 (0.060)	0.215	0.486 (0.048)
2013	0.699 (0.016)	0.353 (0.066)	0.217	0.428 (0.052)
2015	0.676 (0.018)	0.333 (0.076)	0.221	0.408 (0.060)

Notes: Estimates of the mean and variance in neighborhood income during the prior decennial census are used to instrument for each year's values. Composite test scores were standardized with the mean and standard deviation in 1998.

Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. The instrumental variables—the estimated mean and variance of school income—were estimated using the 1980, 1990 and 2000 decennial censuses.

Appendix G: The Relationship Between Student Achievement and Imputed Income

Appendix Table G.1: The Relationship Between Achievement and Alternative Ways of Imputing Income (8th Grade Math)

Year	Instruments Used			
	Race/ Ethnicity	Mother's Education	Urbanicity and State	Race*Moth ed, Urbanicity, State
1992	1.435 (0.024)	0.749 (0.015)	0.805 (0.044)	0.938 (0.015)
1994				
1996	1.333 (0.028)	0.729 (0.016)	0.740 (0.038)	0.938 (0.015)
1998				
2000	1.390 (0.039)	0.721 (0.030)	0.562 (0.044)	0.932 (0.026)
2002				
2003	1.344 (0.017)	0.665 (0.012)	0.493 (0.037)	0.870 (0.010)
2005	1.332 (0.015)	0.690 (0.010)	0.464 (0.029)	0.916 (0.010)
2007	1.237 (0.016)	0.680 (0.010)	0.338 (0.034)	0.827 (0.009)
2009	1.252 (0.017)	0.677 (0.012)	0.422 (0.028)	0.797 (0.010)
2011	1.112 (0.017)	0.640 (0.011)	0.492 (0.027)	0.766 (0.010)
2013	1.131 (0.014)	0.657 (0.010)	0.418 (0.029)	0.776 (0.008)
2015	1.186 (0.018)	0.677 (0.011)	0.517 (0.036)	0.804 (0.009)

Notes: Each column reports the slope of a student-level regression of 8th grade math scores on imputed income, with the imputations based on the variables indicated at the top of each column. Composite test scores were standardized with the mean and standard deviation in 1992.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org. Because the main NAEP assessments are collected between January and March, we use household income from the prior calendar year.

Appendix Table G.2: The Relationship Between Achievement and Alternative Ways of Imputing Income (8th Grade Reading)

Year	Instruments Used			
	Race/ Ethnicity	Mother's Education	Urbanicity and State	Race*Moth ed, Urbanicity, State
1992				
1994				
1996				
1998	1.048 (0.032)	0.606 (0.016)	0.472 (0.046)	0.753 (0.017)
2000				
2002	1.092 (0.024)	0.576 (0.014)	0.296 (0.034)	0.728 (0.014)
2003	1.159 (0.020)	0.637 (0.010)	0.385 (0.034)	0.792 (0.011)
2005	1.165 (0.014)	0.604 (0.008)	0.434 (0.034)	0.800 (0.008)
2007	1.113 (0.015)	0.615 (0.010)	0.239 (0.031)	0.742 (0.009)
2009	1.129 (0.017)	0.607 (0.010)	0.454 (0.027)	0.722 (0.011)
2011	0.998 (0.013)	0.590 (0.011)	0.549 (0.027)	0.694 (0.010)
2013	1.020 (0.016)	0.596 (0.010)	0.432 (0.029)	0.705 (0.009)
2015	1.081 (0.018)	0.597 (0.013)	0.478 (0.032)	0.704 (0.011)

Notes: Each column reports the slope of a student-level regression of 8th grade math scores on imputed income, with the imputations based on the variables indicated at the top of each column. Composite test scores were standardized with the mean and standard deviation in 1998.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Imputed income is based on household income for households with related children aged 5 to 18 and mothers' characteristics in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org. Because the main NAEP assessments are collected between January and March, we use household income from the prior calendar year.

Appendix Table G.3: Between- and Within-School Relationship Between Achievement and Imputed Income (8th Grade Math)

Year	Between-School Slope	Within-School Slope	Student-Level Slope	Number of Students	Number of Schools
1992	1.277 (0.023)	0.667 (0.013)	0.938 (0.015)	75,930	2,900
1994					
1996	1.293 (0.024)	0.644 (0.013)	0.938 (0.015)	67,870	2,920
1998					
2000	1.404 (0.031)	0.609 (0.018)	0.932 (0.026)	60,180	2,610
2002					
2003	1.490 (0.021)	0.602 (0.011)	0.870 (0.010)	103,780	4,070
2005	1.446 (0.020)	0.639 (0.009)	0.916 (0.010)	114,810	4,900
2007	1.271 (0.019)	0.584 (0.009)	0.827 (0.009)	106,750	4,700
2009	1.207 (0.018)	0.579 (0.010)	0.797 (0.010)	111,270	4,740
2011	1.186 (0.016)	0.549 (0.010)	0.766 (0.010)	116,390	4,860
2013	1.185 (0.017)	0.558 (0.010)	0.776 (0.008)	117,150	4,620
2015	1.220 (0.018)	0.548 (0.011)	0.804 (0.009)	97,060	4,310

Notes: Household income is imputed based on student-level characteristics found in the NAEP: race/ethnicity, mother's education, urbanicity, and state. Composite test scores were standardized with the mean and standard deviation in 1992.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Imputed income is based on household income for households with related children aged 5 to 18 in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table G.4: Between- and Within-School Relationship Between Achievement and Imputed Income (8th Grade Reading)

Year	Between-School Slope	Within-School Slope	Student-Level Slope	Number of Students	Number of Schools
1992					
1994					
1996					
1998	1.029 (0.024)	0.561 (0.013)	0.753 (0.017)	62,270	2,620
2000					
2002	1.128 (0.023)	0.534 (0.012)	0.728 (0.014)	81,350	3,550
2003	1.329 (0.021)	0.545 (0.010)	0.792 (0.011)	108,230	4,280
2005	1.254 (0.020)	0.556 (0.009)	0.800 (0.008)	117,880	5,090
2007	1.180 (0.018)	0.520 (0.009)	0.742 (0.009)	117,510	5,000
2009	1.087 (0.016)	0.509 (0.010)	0.722 (0.011)	116,330	5,030
2011	1.114 (0.015)	0.503 (0.010)	0.694 (0.010)	117,460	5,170
2013	1.123 (0.015)	0.499 (0.009)	0.705 (0.009)	123,650	4,870
2015	1.081 (0.017)	0.476 (0.010)	0.704 (0.011)	101,050	4,530

Notes: Household income is imputed based on student-level characteristics found in the NAEP: race/ethnicity, mother's education, urbanicity, and state. Composite test scores were standardized with the mean and standard deviation in 1998.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Imputed income is based on household income for households with related children aged 5 to 18 in the Annual Social and Economic Supplement of the Current Population Survey in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix H: Inferring the Trend in the Black-White Achievement Difference Conditional on Income

The Black-White gap in achievement, as measured in the NAEP, was closing during the period where Reardon (2011) estimated the income-based gaps were sharply widening. Given that mean Black family income is lower than that of White families, the combination of a rising achievement-income gap and a narrowing Black-White gap implies that racial gaps conditional on income must also have been changing. Below, we infer just how much the Black-White gap *conditional on income* would have had to change to reconcile these two divergent trends.

In the NAEP data, we can estimate the unconditional Black-White gap and the unconditional achievement-income gap. However, without student-level data on income, we cannot directly estimate the racial difference in achievement conditional on income. Thus, we use auxiliary regressions of race on income (and vice versa) from the Current Population Survey data to infer the conditional effects of race implied by the bivariate relationships.

Suppose we had estimates of the bivariate coefficients of achievement on race (β_{1t}) and achievement on income (γ_{1t}) as below:

$$Score_{it} = \beta_{0t} + \beta_{1t}Black_{it} + \varepsilon_{it}$$

$$Score_{it} = \gamma_{0t} + \gamma_{1t}Income_{it} + \vartheta_{it}$$

Now, suppose we were seeking to estimate the conditional effect of race, $\theta_{Black,t}$, and income, $\theta_{Income,t}$, in the equation below:

$$Score_{it} = \theta_0 + \theta_{Income,t}Income_{it} + \theta_{Black,t}Black_{it} + \vartheta_{it}$$

To do so, we need to estimate two additional auxiliary regressions using the CPS data in each year:

$$b_{income,black,t} = \text{Black white difference in income}_{\text{in year } t}$$

$$b_{black,income,t} = \text{Regression of black on income}_{\text{in year } t}$$

Using the standard omitted variable bias relationship, each of the bivariate coefficients are a function of the multivariate coefficients and the auxiliary regressions:

$$(1) \quad \hat{\beta}_{1t} = \hat{\theta}_{Black,t} + b_{income,black,t} \hat{\theta}_{Income,t}$$

$$(2) \quad \hat{\gamma}_{1t} = \hat{\theta}_{Income,t} + b_{black,income,t} \hat{\theta}_{black,t} \Rightarrow \hat{\theta}_{Income,t} = \hat{\gamma}_{1t} - b_{black,income,t} \hat{\theta}_{black,t}$$

Rewriting (2) $\hat{\theta}_{Income,t} = \hat{\gamma}_{1t} - b_{black,income,t} \hat{\theta}_{black,t}$ and substituting into equation (1), it is possible to rewrite $\hat{\beta}_{1t}$ a function of $\hat{\gamma}_{1t}$ (income gap) and $\hat{\theta}_{black,t}$ (the direct effect of race):

$$\hat{\beta}_{1t} = \hat{\theta}_{black,t} (1 - b_{black,income,t} b_{income,black,t}) + \hat{\gamma}_{1t} b_{income,black,t}$$

Rearranging terms, the direct effect of race (conditioning on income), $\hat{\theta}_{black,t}$ can be expressed as a function of the two bivariate coefficients on race and income:

$$\hat{\theta}_{black,t} = \hat{\beta}_{1t} \frac{1}{(1 - b_{black,income,t} b_{income,black,t})} - \hat{\gamma}_{1t} \frac{b_{income,black,t}}{(1 - b_{black,income,t} b_{income,black,t})}$$

(Note that if race and income were uncorrelated, the above simplifies to $\hat{\theta}_{black} = \hat{\beta}_1$, i.e., the bivariate Black-White achievement difference is equal to the Black-White achievement difference, conditional on income).

Thus, if we have $\hat{\beta}_1, \hat{\gamma}_1, b_{black,income}, b_{income,black}$ in each year, we can infer the conditional effect of race each year, $\hat{\theta}_{black,t}$. We use two different series of estimates of the income-achievement relationship: those that we estimated with the NAEP and those that would be implied by a 40 percent increase in the income-achievement relationship as reported by Reardon (2011). We estimate the auxiliary regressions, $b_{black,income}, b_{income,black}$, using the March Current Population Survey in each year, limiting the sample to Black and White households with dependent children between the ages of 5 and 18.

Appendix Figure H.1 reports the trend in the Black-White difference in achievement, conditional on income, implied by our estimates and those that would be implied by a 40 percent increase in the slope coefficient between 1992 and 2015. For instance, in grade 4 math in 1992, the unconditional differences in achievement by race and by income—when combined with our estimated slope with respect to income—imply that Black students scored .316 standard deviations below White students with similar In household income. Using our estimates of the changing slope with respect to income, the implied racial difference, conditional on income, remained largely unchanged at .329 standard deviations by 2015.

By contrast, the red line in Appendix Figure H.1 portrays the conditional effect of race required to reconcile Reardon's finding of a 40 percent increase in the bivariate income coefficient with a narrowing race differential. In 4th grade math, Black students would have to have been scoring .136 standard deviations *higher* than White students of the same income level by 2015.

We see the same pattern in all four grade/subject combinations. The only way that the observed decline in Black-White achievement could be reconciled with a 40 percent increase in income-based achievement differentials would be if the Black-White difference in achievement conditional on income changed sign, from negative to positive—implying that Black students were outscoring White students with the same income levels. In contrast, our finding of a simultaneous narrowing of income-based and race-based achievement gaps would have been consistent with a relatively constant Black-White achievement gap conditional on income. See Appendix Tables H.1 through H.4 for the values of $\hat{\beta}_1$, $\hat{\gamma}_1$, $b_{black,income}$, $b_{income,black}$ used for the projections in each grade/subject.

Appendix Table H.1: Inferring the Black-White Difference in Achievement Conditional on Income (4th Grade Math)

Year	Auxiliary Regressions		B-W Diff. (Uncond.)	Income Slope		B-W Diff. Conditional on Income	
				As estimated	Assuming 1.6 percent annual ↑	As estimated	Assuming 1.6 percent annual ↑
	$b_{black, income}$	$b_{income, black}$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\gamma}_1$ 40%↑	$\hat{\theta}_1$	$\hat{\theta}_1$ 40%↑
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992	-0.147	-0.752	-1.083	0.533	0.533	-0.316	-0.316
1994	-0.147	-0.794					
1996	-0.126	-0.730	-1.076	0.506	0.567	-0.308	-0.240
1998	-0.131	-0.749					
2000	-0.130	-0.685	-0.968	0.424	0.601	-0.262	-0.068
2002	-0.119	-0.698					
2003	-0.129	-0.717	-0.944	0.421	0.627	-0.281	-0.055
2005	-0.118	-0.730	-0.911	0.376	0.644	-0.316	-0.023
2007	-0.126	-0.734	-0.896	0.328	0.661	-0.363	0.004
2009	-0.123	-0.731	-0.894	0.351	0.678	-0.333	0.027
2011	-0.135	-0.779	-0.856	0.329	0.695	-0.378	0.031
2013	-0.123	-0.737	-0.857	0.330	0.712	-0.332	0.089
2015	-0.111	-0.746	-0.810	0.302	0.729	-0.329	0.136

Notes: The auxiliary regressions were estimated by regressing an indicator for race on income and income on race using the Current Population Survey. Column (3) reports the Black-White achievement gap (not conditioning on income) in standard deviation units. Column (4) reports the estimated achievement-income slope from Appendix Table E.1, which used neighborhood incomes for NAEP sampled schools. Column (5) reports the achievement income slope assuming a 1.6 percent annual increase (40 percent over 25 years). Column (6) reports the implied Black-White difference conditional on family income that is implied by the estimates in Appendix Table E.1. Column (7) reports the Black-White conditional difference implied by a 40 percent increase in the income coefficient.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Data on race and income are for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table H.2: Inferring the Black-White Difference in Achievement Conditional on Income (4th Grade Reading)

Year	Auxiliary Regressions <i>b</i> _{black, income} <i>b</i> _{income, black}		B-W Diff. (Uncond.) \hat{B}_i	Income Slope		B-W Diff. Conditional on Income	
				As estimated $\hat{\gamma}_i$	Assuming 1.6 percent annual ↑	As estimated $\hat{\theta}_i$	Assuming 1.6 percent annual ↑
					$\hat{\gamma}_i$ 40%↑		$\hat{\theta}_i$ 40%↑
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1992	-0.147	-0.752	-0.890	0.567	0.567	-0.115	-0.115
1994	-0.147	-0.794	-0.932	0.649	0.585	-0.103	-0.175
1996	-0.126	-0.730			0.603		
1998	-0.131	-0.749	-0.808	0.602	0.621	-0.004	0.018
2000	-0.130	-0.685	-0.813		0.640		0.091
2002	-0.119	-0.698	-0.810	0.441	0.658	-0.136	0.101
2003	-0.129	-0.717	-0.802	0.462	0.667	-0.124	0.102
2005	-0.118	-0.730	-0.795	0.357	0.685	-0.244	0.114
2007	-0.126	-0.734	-0.742	0.302	0.703	-0.267	0.175
2009	-0.123	-0.731	-0.710	0.295	0.721	-0.246	0.223
2011	-0.135	-0.779	-0.688	0.195	0.739	-0.381	0.227
2013	-0.123	-0.737	-0.696	0.177	0.758	-0.369	0.269
2015	-0.111	-0.746	-0.694	0.173	0.776	-0.375	0.282

Notes: The auxiliary regressions were estimated by regressing an indicator for race on income and income on race using the Current Population Survey. Column (3) reports the Black-White achievement gap (not conditioning on income) in standard deviation units. Column (4) reports the estimated achievement-income slope from Appendix Table E.2 which used neighborhood incomes for NAEP sampled schools. Column (5) reports the achievement income slope assuming a 1.6 percent annual increase (40 percent over 25 years). Column (6) reports the implied Black-White difference conditional on family income that is implied by the estimates in Appendix Table E.2. Column (7) reports the Black-White conditional difference implied by a 40 percent increase in the income coefficient.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Data on race and income are for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Table H.3: Inferring the Black-White Difference in Achievement Conditional on Income (8th Grade Math)

Year	Auxiliary Regressions		B-W Diff. (Uncond.)	Income Slope		B-W Diff. Conditional on Income	
				As estimated	Assuming 1.6 percent annual ↑	As estimated	Assuming 1.6 percent annual ↑
	$b_{black, income}$	$b_{income, black}$	\hat{B}_1	$\hat{\gamma}_1$	$\hat{\gamma}_1$ 40%↑	$\hat{\theta}_1$	$\hat{\theta}_1$ 40%↑
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1990	-0.145	-0.735	-0.903				
1992	-0.147	-0.752	-1.088	0.620	0.620	-0.223	-0.223
1996	-0.126	-0.730	-1.063	0.618	0.660	-0.174	-0.128
1998	-0.131	-0.749			0.680		
2000	-0.130	-0.685	-1.034	0.608	0.699	-0.110	-0.010
2002	-0.119	-0.698			0.719		
2003	-0.129	-0.717	-0.960	0.542	0.729	-0.161	0.045
2005	-0.118	-0.730	-0.919	0.569	0.749	-0.111	0.086
2007	-0.126	-0.734	-0.873	0.472	0.769	-0.185	0.142
2009	-0.123	-0.731	-0.871	0.509	0.789	-0.140	0.167
2011	-0.135	-0.779	-0.848	0.412	0.808	-0.277	0.166
2013	-0.123	-0.737	-0.836	0.429	0.828	-0.206	0.233
2015	-0.111	-0.746	-0.846	0.430	0.848	-0.219	0.237

Notes: The auxiliary regressions were estimated by regressing an indicator for race on income and income on race using the Current Population Survey. Column (3) reports the Black-White achievement gap (not conditioning on income) in standard deviation units. Column (4) reports the estimated achievement-income slope from Appendix Table E.3, which used neighborhood incomes for NAEP sampled schools. Column (5) reports the achievement income slope assuming a 1.6 percent annual increase (40 percent over 25 years). Column (6) reports the implied Black-White difference conditional on family income that is implied by the estimates in Appendix Table E.3. Column (7) reports the Black-White conditional difference implied by a 40 percent increase in the income coefficient.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Data on race and income are for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

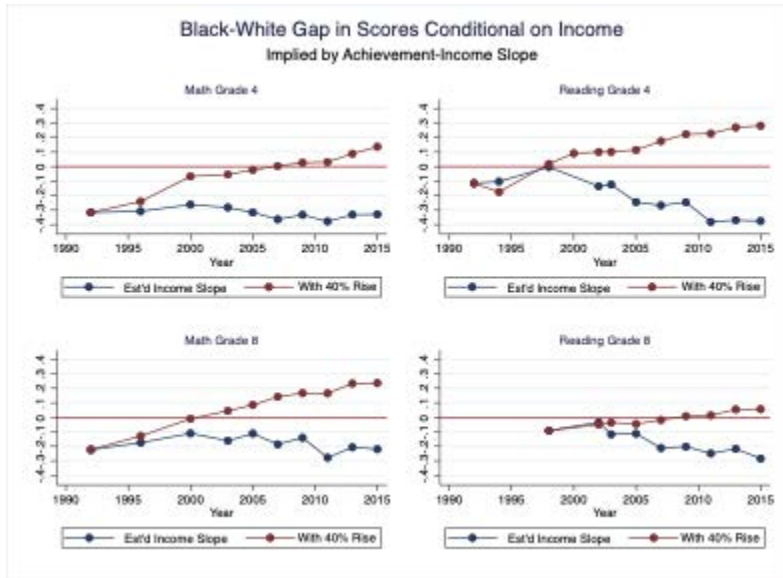
Appendix Table H.4: Inferring the Black-White Difference in Achievement Conditional on Income (8th Grade Reading)

Year	Auxiliary Regressions		B-W Diff. (Uncond.) \hat{B}_1	Income Slope		B-W Diff. Conditional on Income	
	$b_{black, income}$	$b_{income, black}$		As estimated $\hat{\gamma}_1$	Assuming 1.6 percent annual \uparrow $\hat{\gamma}_1$ 40% \uparrow	As estimated $\hat{\theta}_1$	Assuming 1.6 percent annual \uparrow $\hat{\theta}_1$ 40% \uparrow
	(1)	(2)		(4)	(5)	(6)	(7)
1990	-0.145	-0.735					
1992	-0.147	-0.752	-0.813				
1996	-0.126	-0.730					
1998	-0.131	-0.749	-0.744	0.475	0.475	-0.091	-0.091
2000	-0.130	-0.685			0.490		
2002	-0.119	-0.698	-0.789	0.519	0.505	-0.034	-0.049
2003	-0.129	-0.717	-0.761	0.440	0.513	-0.116	-0.036
2005	-0.118	-0.730	-0.780	0.467	0.528	-0.112	-0.045
2007	-0.126	-0.734	-0.763	0.368	0.543	-0.211	-0.018
2009	-0.123	-0.731	-0.753	0.367	0.559	-0.202	0.009
2011	-0.135	-0.779	-0.721	0.339	0.574	-0.249	0.014
2013	-0.123	-0.737	-0.731	0.342	0.589	-0.217	0.055
2015	-0.111	-0.746	-0.741	0.292	0.604	-0.284	0.057

Notes: The auxiliary regressions were estimated by regressing an indicator for race on income and income on race using the Current Population Survey. Column (3) reports the Black-White achievement gap (not conditioning on income) in standard deviation units. Column (4) reports the estimated achievement-income slope from Appendix Table E.4, which used neighborhood incomes for NAEP sampled schools. Column (5) reports the achievement income slope assuming a 1.6 percent annual increase (40 percent over 25 years). Column (6) reports the implied Black-White difference conditional on family income that is implied by the estimates in Appendix Table E.4. Column (7) reports the Black-White conditional difference implied by a 40 percent increase in the income coefficient.

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Data on race and income are for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.

Appendix Figure H.1: Black-White Gap in Scores Conditional on Income Implied by Achievement-Income Slope



Notes: Each point corresponds with the Black-White gap in achievement conditional on income, implied by the bivariate income-race and income-achievement relationships. The blue line uses the estimated income-achievement coefficients from Appendix Tables E.1 through E.4. The red line assumes a 1.6 percent rise in the income-achievement coefficient annually (40 percent over 25 years).

Source: Student test score data are from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Main State sample, for all available years 1992 through 2015. Data on race and income are for households with related children aged 5 to 18 in the ASEC in each year from IPUMS-CPS, University of Minnesota, www.ipums.org.