HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

The Unwavering SES Achievement Gap: Trends in U.S. Student Performance

Faculty Research Working Paper Series

Eric A. Hanushek
Stanford University

Paul E. Peterson
Harvard Kennedy School

Laura M. Talpey
Stanford University

Ludger Woessmann
University of Munich

March 2019

RWP19-012

Visit the **HKS Faculty Research Working Paper Series** at:
https://www.hks.harvard.edu/research-insights/publications?f%5B0%5D=publication_types%3A121

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

The Unwavering SES Achievement Gap: Trends in U.S. Student Performance
Eric A. Hanushek, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann
NBER Working Paper No. 25648
March 2019
JEL No. H4,I2,J24

ABSTRACT

Concerns about the breadth of the U.S. income distribution and limited intergenerational mobility have led to a focus on educational achievement gaps by socio-economic status (SES). Using intertemporally linked assessments from NAEP, TIMSS, and PISA, we trace the achievement of U.S. student cohorts born between 1954 and 2001. Achievement gaps between the top and bottom deciles and the top and bottom quartiles of the SES distribution have been large and remarkably constant for a near half century. These unwavering gaps have not been offset by overall improvements in achievement levels, which have risen at age 14 but remained unchanged at age 17 for the most recent quarter century. The long-term failure of major educational policies to alter SES gaps suggests a need to reconsider standard approaches to mitigating disparities.

Eric A. Hanushek
Hoover Institution
Stanford University
Stanford, CA 94305-6010
and NBER
hanushek@stanford.edu

Paul E. Peterson
Harvard University
Program on Education Policy and Governance
Harvard Kennedy School
Cambridge, MA 02482
USA
ppeterso@gov.harvard.edu

Laura M. Talpey
Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
United States
lmtalpey@stanford.edu

Ludger Woessmann
University of Munich
ifo Institute and CESifo
Poschingerstr. 5
81679 Munich, Germany
woessmann@ifo.de

The increasing disparity in household income and wealth within the United States over the past half century (Krueger (2003); Autor (2014); Saez and Zucman (2016); Alvaredo et al. (2017)) has amplified concerns about the dependency of achievement on a student's socio-economic status (SES). Disparate student outcomes may spell limited intergenerational mobility in the 21st Century. Given the topic's importance, surprisingly little research attention has been given to trends in SES-achievement gaps over the past half century. We draw upon data from four intertemporally linked assessments of student performance administered to representative samples of U.S. students over nearly five decades. Contrary to recent perceptions, we find little change in the SES-achievement relationship across the past fifty years. These gaps occur within the context of stagnant levels of achievement overall, as the steady average gains in student performance registered by age 14 do not lead to gains at age 17 in the most recent quarter century. Our results cast doubt on claims that achievement inequalities are on the rise, but they also suggest a need to reconsider policies and practices pursued in order to ameliorate the SES-achievement connection.

Influential studies have concluded that the broadening dispersion in household income is widening skill gaps between students from advantaged and disadvantaged backgrounds. In *Coming Apart,* Charles Murray (2012) argues that "the United States is stuck with a large and growing lower class that is able to care for itself only sporadically and inconsistently…. [Meanwhile,], the new upper class has continued to prosper as the dollar value of the talents they [sic] bring to the economy has continued to grow." Robert Putnam (2015) says in *Our Kids* that "rich Americans and poor Americans are living, learning, and raising children in increasingly separate and unequal worlds." Richard Rothstein (2004) writes: "Incomes have become more unequally distributed in the United States in the last generation, and this inequality contributes to the academic achievement gap."

The analysts have good reason to express such concerns. The acquisition of cognitive skills early in life is critical for the accumulation of human capital and the amelioration of economic and social inequalities (Friedman (2003); Magnuson and Waldfogel (2008)). Indeed, the U.S. rewards cognitive skills more than almost all other developed countries, which also implies that the U.S. heavily punishes the lack of skills (Hanushek, Schwerdt, Wiederhold, and Woessmann (2015, 2017)). Accordingly, policy makers have long searched for tools that will help schools break the linkage between students' learning and their SES background (Ladd (1996); Carneiro and Heckman (2003); Krueger (2003); Magnuson and Waldfogel (2008)). If the tools applied thus far have been unable to lessen the relationship between SES and achievement, policy makers need to consider alternatives.

The public is accustomed to regular reports of changes in student achievement after release of data from major assessments. However, reporters and analysts typically mention only the most recent changes in achievement levels and gaps (see, for example, Zernicke (2016); Camera (2018)). That focus on the immediate past ignores most of the near fifty-years' worth of data on U.S. student performance in math and reading that has accumulated.

To broaden the perspective, we make use of data from four well-documented and intertemporally linked surveys of achievement in math, reading, and science administered to representative samples of cohorts of U.S. adolescent students who were born between 1954 and 2001. The surveys also provide information on students' SES background.

We find that SES-achievement gaps in the 1950s birth cohorts are very large—hovering around one standard deviation (s.d.) between those in the top and bottom deciles of the SES distribution (90-10 gap) and around 0.8 s.d. for the top and bottom quartiles (75-25 gap). Importantly, these gaps have not increased over time. Instead, they have remained essentially unchanged. That is, students from the most disadvantaged backgrounds have seen trends in achievement similar to those from the most advantaged backgrounds. In terms of learning, students at the 10th SES percentile remain some three to four years behind those in the 90th percentile.

Concern about these gaps would be lessened if all achievement were rising, i.e., if a rising tide was lifting all boats. Test scores for young adolescents (age 13-15) improved consistently over the past fifty years with gains in mean student performance of close to 0.5 s.d. for young adolescents, or roughly 0.1 s.d. per decade. But test scores for older adolescents (age 17) have not kept up. Older adolescents experienced gains in the first half of the period (by 0.1 s.d., or 0.04 s.d. per decade), but in the past twenty-five years their scores have plateaued, suggesting that achievement improvements made by young adolescents are not lasting through their high school years. In other words, the tide does not reach students at the time when they move into college and careers.

This paper has six sections. The first reviews the literature on achievement gaps. The second describes our data, and the third provides our methodological approach. The fourth displays trends in student achievement gaps and levels over the past half century. The fifth considers changes in family and school factors that may have affected the trends. The final section concludes.

1. Relevant Research Literature

Although family background effects on student achievement are well-documented, few studies track changes in the impacts of demographic variables on student performance over time. This lack of longitudinal analysis is partly a function of persistent measurement issues.

1.1 The SES-Achievement Connection

The strong relationship between SES and achievement has long been known (Neff (1938)). Coleman et al. (1966), in their seminal study of *Equality of Educational Opportunity,* found parental education, income, and race to be strongly associated with student achievement, while they concluded school factors to be of much less significance. In a secondary analysis of these data, Smith (1972) also found family background to be the most important determinant of achievement. Subsequent research has confirmed these early findings (Burtless (1996); Mayer (1997); Jencks and Phillips (1998); Magnuson and Waldfogel (2008); Duncan and Murnane (2011); Duncan, Morris, and Rodrigues (2011); Dahl and Lochner (2012); Egalite (2016)).

As discussed elsewhere, many factors connect SES and achievement (Cheng and Peterson (2018)). A few examples illustrate—but certainly do not exhaust—the many possible mechanisms at work. Children exposed to lower SES environments are at greater risk of traumatic stress and other medical problems that can affect brain development (Nelson and Sheridan (2011)). College educated mothers speak more frequently with their infants, use a larger vocabulary when communicating with their toddlers (Hart and Risley (1995, 2003)), and are more likely to use parenting practices that respect the autonomy of a growing child (Hoff (2003); Guryan, Hurst, and Kearney (2008)). College-educated and higher-income families have access to more enriched schooling environments (Altonji and Mansfield (2011)) and are less likely to live in extremely impoverished communities burdened with high violent crime rates

(Burdick-Will et al. (2011)). All these and other childhood or adolescent experiences contribute to profound SES disparities in academic achievement (Kao and Tienda (1998); Perna (2006); Goyette (2008); Jacob and Linkow (2011)).

In empirical analyses, chosen measures of SES are ordinarily based upon data availability rather than conceptual justification. In large-scale assessments of student achievement, data collection procedures usually ignore family-related factors shown to be of importance for student achievement such as parent-child interactions, child upbringing approaches, or general physical and nutritional conditions (see, for example, Gould, Simhon, and Weinberg (2019)). Rather, the general approach is to look for more readily available indicators of persistent cultural and economic differences across families as proxies for the educational input of families. The standard list includes parental education, occupation, earned income, and various items in the home. Age of mother at child's birth, family structure, and a child's number of siblings are also used as indicators of family educational inputs. These measures tend to be highly correlated, making their separate impacts on learning, and their relative importance, difficult to disentangle.

While family income might be thought of as a good summary measure of SES, obtaining data on this from large-scale surveys is problematic. Survey data linked to assessments often come from the students themselves, and students generally have imperfect knowledge of their parents' earned income. For that reason, large-scale assessments that gather information directly from students seek to ascertain economic well-being by asking questions about consumption items, such as the number of durable and educational items present in the home. As compared to household earned income, students are intuitively better informed about whether a durable good (e.g., a dishwasher, computer, or a separate bedroom for themselves) is available in their home (Kayser and Summers (1973); Astone and McLanahan (1991), 313).[1] An analysis by Fetters, Stowe, and Owings (1984) shows that student-reported indicators of parental education tend to be reliable but determined that "family income … was a matter of speculation for many students and thus inaccurately reported" (Kaufman and Rasinski (1991), 2). Consumption indicators may also be useful for estimating the resources of low-income families who supplement earnings with transfer payments, such as food stamps, medical services, housing assistance, and welfare benefits (Slesnick (1993)).

In sum, a child's SES background has been estimated by a variety of measures. The items used depend on alternatives available in the data, but when gathering data directly from students, stable measures such as education and durable goods in the home are generally preferred to annual earnings.

1.2 Changes in the SES-Achievement Connection

A number of scholars have looked at achievement gaps over time. Most studies have traced changes in the size of the black-white test-score gap, but two have explored the evolution of achievement gaps by non-racial SES indicators.

[1] Kayser and Summers (1973) write in their abstract: "In this study, the reliability and validity of student reports of parental SES characteristics was investigated. Using panel data for student reports and independent surveys of both mothers and fathers, it was found that student reports were relatively stable over time and were more reliably measured for parental education than for either father's income or occupation. The validities of the reports were, for all but income reports, moderate. The validity of income reports was very low."

Trends in the black-white gap. Changes in the black-white test-score gap in the United States have been estimated by Grissmer, Kirby, Berends, and Williamson (1994), Grissmer, Flanagan, and Williamson (1998), and Magnuson and Waldfogel (2008). All three rely upon the Long-Term Trend (LTT) version of the National Assessment of Educational Progress (NAEP) to trace changes in the gap for students at ages 9, 13, and 17 between 1971 and the time their studies were completed. All three studies identify a substantial closing of the racial test-score gap for cohorts born between 1954 and 1972. Had the early gains registered by 17-year-olds continued at the same pace in subsequent decades, the black-white test gap would have disappeared for children born in the twenty-first century. But as Magnuson and Waldfogel (2008) put it, "steady gains" occurring among those born just after mid-century "stalled" among cohorts born toward the end of the century, leading to revised projections that it could take more than a century to close the racial gap if progress continues at its more recent pace (Hanushek (2016b)). Reardon (2011) confirms these conclusions.

Since the SES backgrounds of black and white students differ markedly (Magnuson and Waldfogel (2008)), changes in the black-white test-score gap may provide a partial window on trends in the SES-achievement gap. But the correlation between race and SES is declining (Wilson (1987, 2011, 2012)), and since black students constitute only around 16 percent of the school-age population (Rivkin (2016)), this gap cannot provide a complete picture of the SES-achievement gap. It is entirely possible that the mean performance of black and white students could be converging even while SES disparities remain unchanged or increase in magnitude.

Trends in the SES-achievement connection. We are aware of only two studies that provide information on trends in the SES achievement gap.[2] The first of these, by Hedges and Nowell (1998), simply reports observed changes in an Appendix (Table 5B-3, 178). The authors regress student performance on a number of background characteristics, as reported by six nationally representative surveys administered to students between 1965 and 1992. Among the variables included in the regression, the coefficient for parental education is the largest, and it changes little over time. The correlation between achievement and family income in the six surveys is more modest and declines over time.

In a second investigation, Reardon (2011) draws upon data from 12 surveys to estimate gaps in math and reading performances of students at the 90th and the 10th percentile of the household income distribution. In contrast to Hedges and Nowell's results, he finds that the "income achievement gaps among children born in 2001 are roughly 75 percent larger than the estimated gaps among children born in the early 1940s" (p. 95). After 1974, those at the income median were falling further behind those at the 90th percentile. Looking deeper, Reardon (2011) concludes: "The 90/50 gap appears to have grown faster than the 50/10 gap during the 1970s and 1980s" (p. 103). The Reardon study and its conclusions have been widely quoted both by academics and in the general media (e.g., Edsell (2012); Taverise (2012); Weissmann (2012); Maxie (2012); Duncan and Murnane (2014); Putnam (2015); Jackson, Johnson, and Persico (2016)), and the idea that income-achievement gaps have dramatically increased over the past half century may be said to be the contemporary conventional wisdom.

[2] In a somewhat related analysis, Bertrand and Kamenica (2018) consider whether cultural differences defined by income, education, gender, race, and political ideology have widened over time, and they find little evidence of growing cultural divides.

Differences between the findings reported in the two studies may be due to the focus of the first study on overall correlations between SES and mean achievement, while the latter study concentrates on disparities between the extremes of the income distribution. Differences could also reflect the fact that Reardon's analysis makes use of twice as many surveys as the earlier study, including data on more recent cohorts.

We, however, explore a third possibility—the inherent methodological limitations common to both studies. They each estimate trends from data collected by different surveys that are not intertemporally linked but instead are administered to students of varying ages using disparate instruments to estimate achievement levels and SES characteristics. As Eric Nielsen (2015) says, when "data sources have income and achievement measures that do not map easily across surveys, they add an additional layer of complexity and uncertainty to the analysis." It is this uncertainty that we seek to mitigate by relying upon intertemporally linked surveys that allow for consistent measures of both student achievement and SES.

2. Sources of Data

Four surveys use consistent data collection procedures to trace the achievement of representative samples of U.S. adolescents over time. Their tests are designed to be comparable over time by employing psychometric linkage based on using test items that are repeated across test waves. All of them administer low-stakes tests: No consequences to any person or entity are attached to student performances, and results are not identified by name for any school, school district, teacher, or student. All four surveys collect information about the cultural and economic resources of the students' families using student reports of parents' education and of a wide variety of durable material and educational possessions in the home. In addition, parental occupation is available in one survey, and student eligibility for free and reduced lunch is available from administrative records in two of the surveys. Appendix A provides more complete descriptions of the four surveys summarized here.

National Assessment of Educational Progress – Long-Term Trend (LTT-NAEP)

LTT-NAEP tracks performances of adolescent students in math and reading at ages 13 and 17 beginning with the birth cohort born in 1954 who became 17 years of age in 1971.[3] As indicated by its name, this version of the NAEP, often called the "nation's report card," has been developed with the explicit intention of providing reliable measures of student performance over test waves. It is the only source of information for student cohorts born between 1954 and 1976. The U.S. Department of Education suspended administration of the LTT-NAEP in 2014. In a typical year, approximately 17,000 students participate in the administration of the LTT-NAEP.

[3] LTT-NAEP also tests 9-year-olds, but we do not include these data in our analyses in order to maintain a high level of comparability over time as well as to focus on the academic preparation of students as they approach the stage where they need to be career or college ready. For a description of NAEP, see National Center for Education Statistics (2013). In math, the first test is 1973. While we have mean achievement in that year that can be used to analyze trends in achievement levels, we do not have access to the individual student data – making it impossible to calculate the SES gaps for 1973. Thus, the achievement gap analysis is based upon two fewer observations than the level analysis.

Main National Assessment of Educational Progress (Main-NAEP)

Main-NAEP administers tests of math and reading aligned to the curriculum in grade 8.[4] Begun in 1992 with new administrations of the survey every two to four years, it is designed to provide results for representative samples of students in the United States as a whole and for each participating state.[5] Main-NAEP maintains a reputation for reliability and validity similar to LTT-NAEP, and it was thought to track trends over time accurately enough that the LTT-NAEP no longer needed to be administered until 2024. For each administration of the test, the Main-NAEP sample is over 150,000 observations; the large sample is necessary in order to have representative samples for each state.

Programme for International Student Assessment (PISA)

PISA, administered by the Organization for Economic Co-operation and Development (OECD), began in 2000. It was originally designed to provide comparisons among OECD countries, but it has since been expanded to many other jurisdictions. PISA administers assessments in math, reading, and science to representative samples of 15-year-old students every three years. PISA assessments are designed to measure practical applications of knowledge. The United States sample includes over 5,000 students for each administration of the test. The U.S. has participated in every wave of the test, though results are not available for reading for the 1991 birth cohort.

Trends in International Mathematics and Science Survey (TIMSS)

TIMSS, administered by the International Association for the Evaluation of Educational Achievement (IEA), is the current version of an international survey that originated as an exploratory mathematics study conducted in the 1960s in 12 countries.[6] The tests are designed to be curriculum-based and are developed by an IEA-directed international committee. Early IEA tests were not linked over time, but beginning with the cohort born in 1981 (tested in 1995) the TIMSS tests have been designed to generate scores that are comparable over time. We use the TIMSS 8th grade math and science tests beginning with this cohort. The U. S. sample includes approximately 10,000 observations for each administration of the test.

The test years for each of the separate assessments are found in Appendix Table A.1.

3. Methodological Approach

To estimate SES-achievement trends, we aggregate achievement and family background data from the four intertemporally linked surveys and construct an SES index similar to the one used by PISA.

[4] Main-NAEP also tests students in grade 4 and periodically covers a wide variety of other subject areas, none of which are used here. Main-NAEP science is excluded because 8th grade tests were only administered in 2000 and 2005.

[5] Initially 41 states voluntarily participated in the state-representative testing, but the national test results used here are always representative of the U.S. student population. After the introduction of the No Child Left Behind Act of 2001, all states were required to participate in the state-representative tests.

[6] For the history of international testing, see Hanushek and Woessmann (2011).

3.1 Aggregating Data Sets

We compile an aggregate distribution of achievement from student-level microdata available for each subject, testing age, and birth cohort for a fifty-year period. With the exception of 17-year-olds in the LTT-NAEP data, all tests are administered to students between the ages of 13 and 15. The first test was administered by LTT-NAEP in reading to a cohort of students born in 1954; the last test was administered to students born in 2001. Across this near half-century span, achievement data are available for 2,737,583 students from 46 tests in math, 40 in reading, and 12 in science. Table 1 gives for each survey the number of assessments, the subject matter, the age or grade level at which students are tested, the birth cohorts that are surveyed, and the number of observations. Our main sample contains 98 separate test-subject-age/grade-year observations.

The Main-NAEP and TIMSS tests are grade based, while the LTT-NAEP and PISA tests are administered to younger students at slightly different ages. For expositional simplicity, we convert grades to age groups by the modal attendance patterns and refer to all younger students as age 14, the modal age.

To equate results across tests, we calculate achievement means and achievement gaps between groups in s.d. for each subject, testing age, and birth-year cohort. We estimate trends in mean performance over time by calculating the distance (in s.d.) of the mean of the distribution for each test, subject, and cohort observation from the mean score in 2000 (or the closest test year), which is normalized to zero in this base year.[7]

3.2 Constructing the SES Index

We concentrate on two SES-achievement disparities: 1) the difference in achievement between the highest and lowest deciles of the SES distribution (90-10 gap) and 2) the difference between the highest and lowest quartiles (75-25 gap). To do so requires a continuous measure of SES that depicts the full distribution of the population, rather than dividing it into a limited number of categories such as level of degree attainment of parents. Given the inaccuracy of student reports of parental income (and the ensuing lack of such data in large-scale assessments), we construct an index of SES based on student-reported indicators of parental education and home possessions. But we also must take into account changes in the informational content of the indicators over time.

We construct an SES index similar to the one used by PISA (OECD (2017a)), which draws the first principal component from a factor analysis of student-provided data on parental education, parental occupation, and home possessions. We follow PISA by extracting a principal component separately for each test administration, because we aim to observe the percentile distribution of SES in each given year and because the measured home possessions vary over time as does their utility for characterizing SES differences (for details, see Appendix B). We depart from PISA by excluding occupational prestige from our SES index because data on parental occupations are not available from NAEP or TIMSS surveys. Exclusion of the occupational prestige indicator affects the SES index only slightly, because that variable, which estimates occupational prestige by the average education and income of individuals in each occupation, is largely redundant after inclusion of the education and possession variables. The SES index

[7] The base year for all test-subject series is either 1998, 1999, or 2000 with the modal date being 2000.

used here is highly correlated with the PISA index, and the two indices reveal essentially the same trend line in the SES-achievement connection over the period tracked by PISA (Appendix Figure B.2).

Estimating SES by a family's permanent income is conceptually an alternative, but that is not possible from data available in these assessments. Nor is it clear that this is a superior measure of educational inputs of the family. Nonetheless, to judge how our SES indicator correlates with permanent family income, we estimate the correlation between our SES indicator for 1988 and earnings indicators obtained from two waves of a panel survey administered as part of the 1998 Education Longitudinal Study (ELS). Using the average of the two waves as a measure of permanent income, the correlation between individual-level permanent income and our SES indicator is 0.66 (for details, see Appendix B).

3.3 Estimating Trends in Achievement Gaps and Levels

The separate assessments, while internally consistent over time, vary from each other in a variety of details, including relationship to the curriculum, testing philosophy, and sampling frames. We assume that each test is a valid measure of knowledge in each domain even though they vary in content. Differences among tests may also be a function of normal sampling error. To identify the aggregate trend in gaps and levels across birth cohorts, the estimation combines results from all assessments but include indicators for subject, age group, and administrative entity.

For the trends in performance levels, we calculate the mean performance, $\bar{O}^{t}_{isa}$, by subject *s*, testing age *a*, and birth cohort *t* for each survey *i*. We extract the performance trend with a quadratic function of birth year:

$$\bar{O}^{t}_{isa} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \delta_i + \gamma_s + \lambda_a + \varepsilon_{isat} \tag{1}$$

where $\delta_i, \gamma_s, and\, \lambda_a$ are fixed effects for assessment type, subject, and age; *t* is birth year; and ε is a random error. The $\alpha' s$ describe the trend in achievement.

We use the same analytic approach to estimate trends in disparities in student performance for two groups, *j* and *k*, where the gap at any time *t* is Δ^{t}_{jk}. Specifically, we estimate:

$$\Delta^{t}_{jk} = \bar{O}^{t}_{isaj} - \bar{O}^{t}_{isak} = \beta_0 + \beta_1 t + \beta_2 t^2 + \delta_i + \gamma_s + \lambda_a + \mu_{isat} \tag{2}$$

In our main analysis, we estimate these disparity trends for two specific types of gaps:

1. The unconditional gaps between students at different points on the achievement distribution: We show changes in the inter-quartile range as well as the difference between those performing at the 90th and 10th percentiles of the achievement distribution.
2. Disparities by SES background: We consider the achievement gap between those in the top and bottom deciles of the SES index distribution and those in the top and bottom quartiles of the distribution.[8] For expositional purposes, we refer to these as the 90-10 and 75-25 SES gaps.

[8] Note that the gaps for the SES distributions are calculated as the average score for those above the 90th SES percentile and the average score for those below the 10th percentile (and similarly for quartiles). While we can

We supplement these main analyses with two additional analyses:

3. As a robustness check, we estimate gaps in performances between those eligible and not eligible for free and reduced-price lunch.

Students who come from households at or below 130% of the poverty line are eligible for free lunch (who we refer to as extremely poor), while those from households between 130% and 185% of the poverty line are eligible for participation in the reduced-price lunch program (who we refer to as poor). Information on student achievement by eligibility for these federal programs is available for cohorts born as early as 1982. The variable is generated administratively from student records. We compare those eligible for subsidized lunch to those not eligible.

The variable of free or reduced-price lunch eligibility has important limitations. First, it is dichotomous, dividing the distribution at a point near its mean, so it does not allow for estimation near the extremes of the continuum. Second, the share of the population who participate in the free lunch program increases over time for a combination of reasons that include administrative changes in the programmatic rules that allowed new eligibility certification and allowed entire schools to participate in the program. For example, comparing academic year 1999 and 2015, the percentage of children below 200 percent of poverty was virtually identical (39 percent), but the percentage in the free and reduced price lunch program increased from 37% to 52% (Chingos (2016); Greenberg (2018)). For these reasons, we regard this variable as only a crude SES indicator that is best used as a robustness check.

4. To allow for comparisons between SES-achievement and race-achievement gaps, we also estimate the black-white test-score gap with NAEP data.

In terms of racial differences, both the LTT-NAEP and Main-NAEP use school-district administrative data to classify students by their racial and ethnic background. PISA does not collect race information in a comparable form, and TIMSS, which collects information on race from student questionnaires, does so for only a subset of its survey administrations. We do not track disparities for other ethnic groups. Continuous immigration has substantially altered the composition of Asian and Hispanic populations over the past 50 years, complicating comparisons of test performance for these groups over time. However, we do estimate the SES-achievement gap separately for white students.

4. Trends in Achievement Gaps

Our results indicate that achievement gaps have been wide and persistent for the last half century. We begin with the aggregate trend in the SES-achievement gap for all students in all subjects and then explore heterogeneities by subject, ethnic group, and an alternative measure of income. The persistent gaps observed might be less disconcerting if achievement levels were rising for everybody, making the economic future better across the SES spectrum. But while we find steady achievement gains across

calculate the precise 90th and 10th percentile values for the distribution of our SES index, this does not correspond to a specific individual or specific test score. Thus, we average test scores across all students in the relevant tails of the SES distribution. For the unconditional gaps, the 90th and 10th percentiles are specific scores in the achievement distribution, and we use these values to calculate gaps and not the average performance in the tails.

cohorts for younger students, these gains do not carry forward to age 17, the time when students are leaving the secondary schooling system.

4.1 Unconditional Achievement Disparities

We begin by estimating changes in the overall distribution of achievement. At the top of Figure 1, we plot the unconditional gaps measured in initial standard deviations for the 90-10 and the 75-25 gaps over the past half century. The nonlinear trend estimates are based on Equation (2) where trends are extracted by taking a quadratic function of the birth year. The gap between those at the 90th and 10th percentile of the achievement distribution among those born in 1954 is close to 2.4 s.d.[9] Over the next fifty years, this gap (measured in units of the initial s.d.'s) closes slightly to 2.16 s.d., indicating some shrinkage in the overall variance of achievement.

The unconditional 75-25 gap, or inter-quartile range, in the achievement distribution is, by definition, smaller than the 90-10 gap. For students born in 1954 it is 1.3 s.d. Over the next fifty years, the inter-quartile range declines modestly by 0.15 s.d. In sum, the overall distribution of achievement, while narrowing a little, has shown only limited change. Students at the bottom of the achievement distribution have seen the same (or slightly more favorable) change in achievement as those at the top.

Looking at results by subject, in math both the 90-10 gap and the 75-25 gap close somewhat over the first half of the observation period but remain mostly flat at the end of the period (not shown). Gaps in reading are even more constant, with a very slight tendency to increase initially and a slightly smaller tendency to fall over the second half of the observation period (not shown). We also find no difference by age group (not shown).

4.2 Achievement Disparities by SES

The pattern of the trends in SES-achievement gaps in Figure 1 is startling: The connection between SES and achievement hardly wavers over this half century. In the 1954 birth cohort, the achievement gap between the average of those in the top and bottom deciles of the SES distribution stood at slightly less than 1.2 s.d.[10] For those born in 2001, the gap is slightly less—about 1.05 s.d. That is, the most disadvantaged students in terms of SES background have seen essentially the same change in achievement as the most advantaged students.

The gap between students in the top and bottom quartiles of the SES distribution was about 0.9 s.d. for the 1954 birth cohort. As the trend line in Figure 1 indicates, this gap declines to barely below 0.8 s.d. for the cohort born in 2001.

Trends are quite similar for math and reading separately. The gap in math achievement, particularly for the 90-10 comparison, shows a little movement over the period—narrowing in the early years but returning to a position below the initial level in recent decades (Figure 2a). The 75-25 math gap narrows slightly over time. In reading, the pattern appears essentially flat for the entire period (Figure 2b).

---

[9] If measured performances were normally distributed, the 90-10 gap would be 2.56 s.d., but the test score distribution is obviously truncated at the extremes.

[10] As noted above, the unconditional gaps and SES gaps are calculated differently, with the 90-10 SES gaps indicating differences between the average students in the top and bottom ten percent of the SES distribution.

Figures in Appendix C display plots of unconditional gaps and SES gaps for each of the individual assessments. Generally speaking, the trend lines for gaps within each assessment resemble those for the aggregate trend lines reported above. The plots for the LTT-NAEP 13-year-old and 17-year-old scores resemble one another, and both show very little fluctuation over time (Appendix Figure C.1). The Main-NAEP math assessments (Appendix Figure C.2A) and the TIMSS assessments (Appendix Figure C.3) show slight increases in the 90-10 SES gaps but not in the 75-25 SES gaps. On the other side, the PISA math, reading, and science assessments (Appendix Figure C.4) all show some narrowing of the SES gaps for the 1985 birth cohorts and later. Overall, the four underlying assessments produce patterns of the SES gaps that are all similar—and yield the unwavering trends of Figure 1.[11]

Our findings confirm Reardon's (2011) identification of large gaps in academic performance between students at the extremes of the SES distribution,[12] but we are unable to replicate his finding that achievement differentials have risen by as much as 75 percent over the past fifty years. His results may be a function of a reliance upon cross-sectional studies that use disparate methods for collecting both income and achievement information. Whatever the reason, the size and trends estimated there differ markedly from the trend in SES-achievement gaps estimated from intertemporally linked surveys administered consistently over time.

In sum, any increase in the disparities in wealth, earnings, and income that may have occurred over the past half century do not translate into an increased connection between students' family backgrounds and their achievement levels in adolescence. Instead, all the trend lines in SES-achievement disparities are basically unchanged with no indication of any long-term upward trajectory.

4.3 Additional Analyses of Achievement Disparities

These findings are confirmed by estimations based on student eligibility for free and reduced-price lunch, on racial groups, and on data adjusted for changes in the ethnic composition of the population.

Eligibility for free and reduced-price lunch program. As a robustness check, we estimate the gap between students who are eligible and those who are not eligible for participation in the federal free lunch program at school.[13] As can be seen in Figure 3, the gap between the extremely poor students and other students in the 1982 birth cohort is a sizeable 0.71 s.d. When the extremely poor are combined with the poor, the gap for this cohort is nearly as large – still 0.64 s.d. Over the next twenty years, the gap between the extremely poor and students from families above the eligibility line narrows

[11] Note that LTT-NAEP for 13-year-olds, Main-NAEP, and TIMSS show some increase in the 90-10 SES gap since the mid-1970s, while PISA shows steady declines since the 1985 birth cohort (the first observed). As a sensitivity analysis, we drop the PISA tests from the trend estimation. As shown in Appendix Figure D.1, this produces somewhat more bow in the trend line of the 90-10 SES gap (but not the 75-25 SES gap). For the equivalent of Figure 1 without PISA, the 90-10 gap starts at 1.25 s.d., falls to 0.95 s.d. for the 1980 birth cohort, and rises back to 1.16 s.d. for the 2001 birth cohort. We do not see a reason, however, why the PISA data should be less valid than the other data.

[12] From the figures in Reardon (2011), we estimate an average 90-10 income gap of close to one s.d., virtually the same as the average 1.03 s.d. gap that we identify for the 90-10 SES disparity.

[13] The analysis of free and reduced-price lunch eligibility relates solely to assessments from Main-NAEP, the only survey to include such information.

by 0.06 s.d. and the gap between ineligible students and all those eligible for participation in the program narrows by 0.01 s.d.

Just like the results based on the SES index, this measure of the earnings-achievement gap reveals only miniscule change over the course of two decades. These results are entirely consistent with the trends for both the 75-25 and the 90-10 SES-achievement gaps reported above.

Achievement by racial group. To facilitate a comparison of trends in the SES-achievement and race-achievement gaps, we also report the black-white test-score gap in Figure 3. Our results confirm – and update to a more recent period – what other scholars have shown. The black-white gap declines from about 1.3 s.d. for the 1954 cohort to about 0.8 s.d. for those born thirty years later – a closing of greater than 0.1 s.d. per decade. But the gains do not continue to accumulate after that point. This stalled progress pointed out by Magnuson and Waldfogel (2008) is consistent with the evidence in Reardon (2011) that shows a decline of about 0.5 standard deviations in the black-white test-score gap in reading for cohorts born between 1950 and 1980 and a slower subsequent rate of change.

Clearly, efforts to close the racial achievement gap in the United States have been more successful than endeavors to close the SES-achievement divide, at least until about 20 years ago. For the past two decades of student cohorts, both the race-achievement gap and the SES-achievement gap have remained essentially flat.

Ethnic composition. Some have hypothesized that the lack of success in diminishing the size of the SES gap is due to changes in the racial and ethnic composition of the school population, as the ethnic make-up of the U.S. population has changed dramatically over the past half century. In 1980, the population age 5-17 was 74.6 percent white, 14.5 percent black, 8.5 percent Hispanic, and 2.5 percent other. In 2011, the corresponding figures were 54.2 percent white, 14.0 percent black, 22.8 percent Hispanic, and 8.9 percent other.[14]

To see whether trends in achievement gaps are driven by shifts in ethnic composition, we estimate the SES-achievement gap just for white students (Figure 4). As expected, the unconditional 90-10 and 75-25 gaps are slightly smaller than in the full population. The gaps have declined slightly over the first half of the observation period and flattened out since.

Turning to the SES divide, the substantively meaningful 90-10 SES-achievement gap for the white cohort born in 1954 was 0.97 s.d. By the middle of the period, the divide had declined by about 0.3 s.d., but it then rose by a commensurate amount so that 90-10 SES gap for the 2001 white birth cohort is just 0.1 s.d. smaller than the gap for the 1954 cohort.

The 75-25 SES-achievement gap among whites stood at about 0.8 s.d. for the 1954 birth cohort. It eased by about 0.25 s.d. toward the middle of the period only to return to just under its original level by the end. In other words, the minor fluctuations for whites parallel the minimal ones observed for all students (Figure 1), supporting the conclusion that changes in the ethnic composition of student cohorts do not account for the unwavering SES-achievement gap.

[14] The large jump in the "other" category includes a substantial jump in the Asian population (to 4.4 percent) and the addition of 4.6 percent identified as two or more races—a category that was not reported in 1980 (U.S. Department of Education (2013), Table 20).

4.4 Achievement Levels

The disappointing lack of improvement in the distributional patterns might be less of a concern if they were offset by improvements in the overall level of achievement. Using the time series data on student outcomes, we can directly evaluate whether there are any gains in student achievement and, importantly, whether they persist until the completion of secondary schooling.

Figure 5 shows a significant upward trend in the overall mean achievement level of adolescent students of approximately 0.3 s.d. over the course of the last half century, or approximately 0.06 s.d. per decade. The nonlinear trend estimates based on Equation (1) are again extracted by taking a quadratic function of the birth year. Importantly, the gains are concentrated on the performances of adolescents who are age 14 or less, where an overall increase of about 0.46 s.d. is observed, approximately 0.09 s.d. per decade. By contrast, gains among students at the age of 17 are only about 0.1 s.d., and no gains are observed for older students after the 1970 birth cohort. The rising tide of student achievement does not extend to students on the cusp of moving into careers and college.

The average improvement seen in test performance among those at age 14 (LTT-NAEP, Main-NAEP, and TIMSS) are larger than those registered in the PISA tests, which are administered at age 15 (not shown).[15] This may be due to differences in test design or it may suggest that the aggregate score fade out begins in the early years of high school.[16]

Nonetheless, it is natural to expect gains realized by ages 13 to 15 to remain intact or even grow by age 17. We return to this puzzle below, although we have no easy answer for this break in learning gains.

There are significant heterogeneities in the trends in achievement level by subject. Mean achievement gains by cohorts are largely concentrated in mathematics. Younger adolescents register a math improvement of 0.9 s.d., while the older ones show an overall shift upward of 0.2 s.d. (Figure 6a). Reading gains are smaller. The trend among older adolescents shows no improvement, while the trend among younger adolescents amounts to only 0.23 s.d. over the half century (Figure 6b).[17] These subject differences are consistent with a general finding that schools and teachers appear to have a significantly stronger impact on math than on reading, something generally attributed to lesser parental influence on math learning (e.g., Hanushek and Rivkin (2010)).

[15] The performance levels of 17-year-old students are not significantly affected by changes in ethnic composition discussed earlier. To see this, it is possible to estimate the LTT-NAEP scores for 2012 if the population had the same ethnic distribution as in 1980. In particular, we can weight the 2012 math and reading scores of white, black, Hispanic, and other groups by the 1980 population distribution of these groups. The estimated 2012 math score for 17-year-olds is 309 versus the actual score of 306, or a difference of 0.08 s.d. over the entire period. For reading, the estimated score with 1980 weights is 289 versus the actual score of 287, or a difference of 0.07 s.d. over the time period.

[16] Two-thirds of PISA students are in grade 10 with the remainder roughly evenly divided between grades 9 and 11.

[17] We do not analyze science separately because the number of science observations is limited and the period is shorter than for the other two subjects.

Importantly, the trends in achievement gaps considered earlier are essentially the same for both the younger and older students (not shown). In both cases, we detect very little temporal change in achievement gaps.

4.5 Summary of Results

Performance disparities are both large and extraordinarily persistent. The SES-achievement gap within the United States has remained essentially as large as in 1966 when James Coleman wrote his report on *Equality of Educational Opportunity* and the United States launched a national "war on poverty" in which compensatory education was the centerpiece. In terms of learning, students at the 90^{th} percentile of the SES distribution are three to four years ahead of those at the 10^{th} percentile by 8^{th} grade. These SES-achievement gaps are amazingly large and unwavering.

Nor does this constancy reflect broad success where all SES groups have benefitted from improved outcomes over time. Students in their early adolescent years show achievement gains over the past half century. Students appear better prepared for entry into high school than they were five decades earlier. The gain is about 0.46 s.d. or about 0.09 s.d. per decade. The achievement gains of young U.S. adolescents are comparable to average gains in other countries that have tracked student progress over time. Between 1995 and 2009, the PISA and TIMSS test score performances of elementary students and young adolescents in 47 countries improved in math, reading and science by 0.12 s.d. per decade (Hanushek, Peterson, and Woessmann (2012)), somewhat larger than the 0.08 s.d. gains per decade over the near 50-year period reported above.

But these gains lead to a puzzle: Over the past quarter century, achievement gains apparent among students at age 14 disappear by the age of 17. As students reach the point of entering college or the labor market, advances in performance are no longer seen.

5. Discussion

Despite the paucity of trend studies, static consideration of achievement gaps and gains has long been a topic for systematic research and policy discussion. A simple educational production function model underlies much of the public and academic discourse, namely:

$$achievement = family\ inputs + schools + other \qquad (3)$$

Our aggregate trend data cannot of course identify the causal effect of each component of this relationship, but it is possible to see if performance trends are consistent with either demographic changes or policy shifts, or both. To generate hypotheses for future research, we therefore summarize major changes in family inputs and school policies to see whether one or another (or both) might plausibly account for the trends.

A substantial body of existing research on student achievement attempts to parse components of schools ($S_1, S_2, S_3, \ldots$) that are causally related to student outcomes, and some, albeit fewer, pursue estimates of the impacts of family components ($F_1, F_2, F_3, \ldots$). From these studies, one can at times obtain credibly estimated relationships for individual factors under specific circumstances – generally from micro-level studies of student achievement. However, there are few if any attempts to consider the aggregate impacts of these demographic shifts and policy thrusts.

5.1 Assessing Trends in Achievement Levels

To open further inquiry into this topic, we discuss here what can be learned from the demographic changes and policy innovations that have taken place over the past half century. Conceptually, we difference Equation 1 between two time points (t_0 and t_1) and write a linear version in terms of one family or school input factor (X_l):

$$\bar{O}_{isa}^{t_1} - \bar{O}_{isa}^{t_0} = \alpha_{X_l}(X_l^{t_1} - X_l^{t_0}) + \eta_{isa} \quad (4)$$

As we have discussed, Figure 5 shows gains in average achievement for 14-year-olds but not for 17-year-olds.[18] We do not have a full set of impact parameters (α_{X_l}), so it is not possible to solve for the relative impact of various demographic and school factors. Nonetheless, many presume that the gains for young adolescents are plausibly a function of positive changes in family demographic characteristics known to be correlated with educational performance, such as parental education, household income, family size, and age of mother at child's birth (Hoxby (2003); Magnuson and Waldfogel (2008)).

Grissmer, Kirby, Berends, and Williamson (1994) correlate shifts in a battery of demographic factors with changes in student performance on the LTT-NAEP between 1970 and 1990. Their estimate (p. 92) shows that changes in family background characteristics can account for all of the math gains in student achievement among students at age 13 and 17. Changes in family background actually over-predict reading gains. Similarly, Duncan, Kalil, and Ziol-Guest (2017), drawing upon longitudinal data available from the Panel Survey of Income Dynamics (PSID), identify income, education, number of siblings, and age of the mother as positive factors affecting years of education and college completion rates. According to this analysis shifts in family background factors can account for all the gains in student performances among young adults over the past half century.[19] That the gains in the United States are comparable to the average gains registered on international tests elsewhere in the industrialized world adds weight to this interpretation (Hanushek, Peterson, and Woessmann (2012)).

The puzzling disappearance of achievement gains by age 17 that has occurred over the past quarter century makes the trends more difficult to interpret. It cannot easily be attributed to family background factors, because one assumes that family cultural and economic resources to be no less important for the performances of older students than they are for those in eighth grade.

While this difference between trends for younger and older students has been noted previously, no satisfactory explanations have been found (Krueger (1998); Hanushek (1998)). Blagg and Chingos (2016) consider four potential reasons for the fade out but reject each. The decline does not appear to be due to increases in the share of the cohort in school at age 17, because trends in performance are uncorrelated with trends in graduation rates. Nor do they attribute it to changes in ethnicity and other

---

[18] Note that the PISA observations for 15-year-olds are included in the 14-year-old plot.

[19] Note, however, that neither of these analyses includes any measures of schools, so that the demographic factors implicitly include correlated differences in school quality.

family background characteristics, because the differential trends persist even when adjusted for demographic changes. Nor do they think it could be a function of a decoupling of the LTT-NAEP from the high-school curriculum, as fade out is also apparent on the Main-NAEP, which has been designed to test performance on material that is part of the curriculum.[20] Nor does it seem to be a function of changes in "senioritis," the propensity of 17-year-olds to take tests less seriously than younger students, as they find no change over time in the number of unanswered questions and other indicators of disengagement.

We offer two hypotheses for the perplexing discrepancy between achievement trends for younger and older students. First, the changes in factors likely to affect teacher quality (discussed further below) may be more unfavorable for instruction at the upper secondary level than at the elementary and middle school levels. Teacher salaries have become more compressed (Hoxby and Leigh (2004)), a likely indication that secondary teacher salaries have declined relative to those earned by elementary teachers. Second, policy initiatives have focused primarily on elementary and middle school. School accountability under No Child Left Behind, for example, required testing each year in grades three through eight but only required one examination in high school. Nonetheless, given available evidence, it is difficult to assess the relative importance of each factor.

5.2 Assessing Trends in Achievement Gaps

When we turn to the story on achievement gaps, we find an even larger puzzle. In terms of changing achievement disparities across SES groups, the key is the relative inputs received by SES group. We define $\lambda^{t}_{X_l jk}$ as the difference in input X_l received by group *j* relative to group *k* in year *t*. Then, parallel to the level patterns, the trend in achievement gaps for *j* relative to *k* is simply a function of the change in relative resources of the two groups:

$$\Delta^{t_1}_{jk} - \Delta^{t_0}_{jk} = \beta_{X_l}(\lambda^{t_1}_{X_l jk} - \lambda^{t_0}_{X_l jk}) + \omega_{ias} \tag{5}$$

where X_l is the given school or family input. If we knew the β_X, say from appropriate micro-studies, we could assess the importance of changes in various school inputs.

When we look at the patterns of SES achievement gaps, however, we see that in general there is not a time trend:

$$\Delta^{t_1}_{jk} - \Delta^{t_0}_{jk} = 0 \tag{6}$$

Interpreting the pattern in the SES-achievement gap is challenging when no clear trend—either upward or downward—is detected. But, as Jencks and Phillips (1998), p. 27, observe in their classic collection on the black-white test score gap, "we have to explain stability as well as change."

[20] Note that 17-year-old assessments are found only in the LTT-NAEP sample, and these assessments are designed to test a consistent body of material over time.

We can consider this as a general problem of reverse engineering in the policy space. In particular, we do not have reliable estimates for the β_X that apply to most major policy initiatives (as discussed below), but from Equations 5 and 6 we know that

$$0 = \beta_{X_l}(\lambda^{t_1}_{X_l jk} - \lambda^{t_0}_{X_l jk}) + \omega_{ias} \tag{7}$$

Two hypotheses are worthy of consideration. The first is the simple null hypothesis, which attributes the lack of any trend to the absence of a significant causal factor that might have closed or widened the gap, i.e., $\beta_X = 0$, along with the expected value of the error term being zero, $E(\omega_{ias}) = 0$. An alternative to the null hypothesis is one that identifies equally powerful but opposing forces which cancel one another out, i.e., $E(\omega_{ias}) \neq 0$ because of one or more other correlated factors. For example, changes in society may have aggravated achievement inequalities, while schools have offset their impact. Or vice versa.

Reardon (2011) attributes the rising achievement gap he observes to the widening differential in household income. It is also possible that SES differences in the age of the mother at the birth of the child that have opened up in the past fifty years enter negatively (Duncan, Kalil, and Ziol-Guest (2017)). And, the incidence of single-parent households is concentrated at the lower end of the SES spectrum. But all these negative factors could be offset by other, countervailing demographic changes. Most importantly, SES differences in parental education have narrowed. So have SES differences in the number of siblings in the household. Both factors have been identified as among the most important determinants of student achievement. The balance between these countervailing factors may well have left the achievement gap pretty much at the same level today as it was for cohorts born in the 1950s.

As for schools, following the null hypothesis, one might conclude that little of significance for achievement has changed. The organization of the country's K-12 education system is much the same in the 21st Century as it was in the middle of the 20th Century. Schools are still operated by relatively autonomous school districts under the control of (mainly) elected school boards. The length of the school year has not changed. Teacher recruitment policies remain substantially the same, and teachers are still compensated according to a standardized salary schedule that rewards experience and academic credentials, not classroom effectiveness.

Yet there have been a set of major policies designed to close achievement gaps. Since 1960 – the year the earliest cohort entered school – a variety of significant policy changes have been adopted as a way of meeting the needs of disadvantaged students; i.e., $(\lambda^{t_2}_{S_l jk} - \lambda^{t_1}_{S_l jk}) > 0$:

- The 1954 Supreme Court decision in *Brown v. Board of Education* led to substantial school desegregation particularly in the South (Welch and Light (1987); Rivkin and Welch (2006); Rivkin (2016)).
- With the advent of the war on poverty, the Title I of the Education and Secondary Education Act (ESEA) of 1965 directed federal compensatory education resources to school districts with disproportionately large shares of low-income students, though a portion might have been offset by reduced state and local funding (Cross (2014)).

- In 1974, the Education for All Handicapped Children Act, later renamed the Individuals with Disabilities Education Act, authorized grants to school districts with accompanying restrictions that assured the provision of educational and other services to those with disabilities, a group disproportionately comprised of students from low-income families (Morgan, Farkas, Hillemeier, and Maczuga (2017)).
- States systematically changed their funding of local schools, often in response to court orders requiring greater fiscal equality among school districts. (For a history and discussion of school finance litigation, see Peterson and West (2007) and Hanushek and Lindseth (2009); see also Jackson, Johnson, and Persico (2016) and Lafortune, Rothstein, and Schanzenbach (2018).) These changes led to more funding equality between districts serving the most disadvantaged and those serving the least disadvantaged.
- The federal Head Start program and expanded state programming provided new opportunities for early childhood education for low-income families (Friedman-Krauss et al. (2018)).
- Accountability for student performance was introduced, first by individual states and then nationally with the enactment in 2002 of the No Child Left Behind Act. The law's accountability requirements were disproportionately directed toward schools serving low-income students (Hanushek and Raymond (2005); Peterson (2010); Figlio and Loeb (2011)).

Of these items, school desegregation is noteworthy for its bivariate credibility. In the aftermath of *Brown*, it took considerable time before schools were substantially desegregated, but by the 1970s and 1980s there was noticeable desegregation of schools, which is also the time when the black-white test score gap closes. After 1980, the rate of desegregation slows to a near stop, and so does the closing of the black-white gap (Figure 3). Micro-studies generally find that school desegregation boosts black achievement, though the evidence is mixed.[21]

However, it is not clear whether school desegregation mitigated SES-achievement disparities. The achievement gains from desegregation may have been disproportionately concentrated on black students from high SES families. If so, its contribution to the closing of the SES gap remain uncertain.

Other changes in school policy may have had positive effects on low-SES students. Overall school funding increased dramatically on a per pupil basis, quadrupling in real dollars between 1960 and 2015. A large portion of this spending increase went toward reductions in pupil-teacher ratios (see Hanushek and Rivkin (1997)). While we do not have clear evidence as to whether these policies disproportionately affected low-SES students, a large share of the monies was directed toward central-city school districts.

Alternatively, programs that advantage neither group may affect achievement gaps if the β_S differ across the groups. Jackson, Johnson, and Persico (2016) find that resource increments have larger impacts on the performances of disadvantaged students, implying that the increases in overall funding induced by court orders have led to a significant reduction in achievement gaps. However, they express doubt as to whether their results generalize to most spending increases.

[21] There is substantial causal evidence from micro-studies that indicates a positive impact on black achievement with desegregation (see Hanushek, Kain, and Rivkin (2009)), but there is earlier mixed evidence (Schofield (1995)).

It is nonetheless more likely that the unwavering gap is due to counter-vailing forces within the educational system that are offsetting one another. Most significantly, the quality of the teaching force—a centrally important school input affecting student achievement—may well have declined over the course of the past several decades.[22] Women have greater access to opportunities outside the field of teaching (Eide, Goldhaber, and Brewer (2004); Corcoran, Evans, and Schwab (2004a, 2004b); Bacolod (2007)). Teacher performance on standardized tests and indicators of teacher selectivity have slipped (Corcoran, Evans, and Schwab (2004a); Bacolod (2007)). Teacher salaries have declined relative to those earned by other four-year college degree holders (Corcoran, Evans, and Schwab (2004a); Hoxby and Leigh (2004); Hanushek (2016a)), and salary levels for teachers are currently low relative to comparable workers in other occupations (Hanushek, Piopiunik, and Wiederhold (forthcoming)).

These changes affecting the quality of the teaching force are likely to have had a disproportionately adverse effect on disadvantaged students. More experienced teachers have acquired seniority rights, and new entrants into the labor force are assigned to more disadvantaged students (Hanushek, Kain, and Rivkin (2004); Loeb, Kalogrides, and Béteille (2012); Kalogrides, Loeb, and Béteille (2013)).

The flat pattern of achievement gaps suggests that the combined positive impact of all of the major policies has not been sufficient to offset any decline in teacher quality.

5.3 Summary

The SES-achievement gap may persist because changes within families and within schools have largely offset one another. When it comes to family background, reduced disparities in family education and family size could have been counterbalanced by rising gaps in family structure, age of the mother, and household income. When it comes to schools, compensatory education policies at both pre-school and K-12 levels may have been offset by an inability to prevent rising gaps in teacher quality across the SES spectrum.

Without adequate estimates of the causal impact of the various changing inputs, it is not possible to reach firm conclusions about impacts and offsets of existing demographic patterns and policy initiatives. But the overall patterns of performance give little confidence that current policies are having much overall influence on improving social outcomes in terms of the level or distribution of achievement.

6. Conclusions

Two startling results emerge from this analysis of long-term trends in student achievement gaps and levels across the SES distribution. First, gaps in achievement between low and high SES groups are mostly unchanged over the past half century. Second, while gains in the level of achievement are steady and significant at the 8th grade level, they have not translated into gains at the end of high school. Thus, the continuing unequal opportunities of the haves and the have nots are not compensated for by enhanced overall opportunities.

Because cognitive skills as measured by standard achievement tests are a strong predictor of future income and economic well-being, the unwavering achievement gaps across the SES spectrum do not

[22] For discussions of the measurement of teacher quality and its importance in school performance, see Hanushek and Rivkin (2006) and Chetty, Friedman, and Rockoff (2013).

bode well for improvements in intergeneration mobility in the future. Perhaps more disturbingly, the U.S. has introduced and expanded a set of programs designed to lessen achievement gaps through improving the education of disadvantaged students, but they individually and collectively appear able to do little to close gaps beyond offsetting the probable decline in teacher quality in schools serving lower SES students. These unwavering gaps suggest reconsidering existing policy thrusts.

Two areas for further policy exploration seem especially critical. First, researchers have uniformly found that teacher effectiveness is a predominant factor affecting school quality. While there has been considerable public discussion of teacher evaluations and programs aimed at teacher effectiveness, few programs and policies at scale directly focus on enhancing this resource, particularly for disadvantaged students. Second, the trend line for those in their final year of high school is much less favorable than for students at an earlier age. Yet most policy interventions have left high schools essentially untouched.

References

Altonji, Joseph G., and Richard K. Mansfield. 2011. The role of family, school, and community characteristics in inequality in education and labor-market outcomes. In *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, edited by Richard J. Murnane and Greg J. Duncan. New York: Russell Sage Foundation Press: 339-358.

Alvaredo, Facundo, Lucas Chanel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, eds. 2017. *World Inequality Report 2018*. Paris: World Inequality Lab.

Astone, Nan Marie, and Sara S. McLanahan. 1991. "Family Structure, Parental Practices and High School Completion." *American Sociological Review* 56, no. 3 (June): 309-320.

Autor, David H. 2014. "Skills, education, and the rise of earnings inequality among the “other 99 percent”." *Science* 344, no. 843 (May 23): 843-851.

Bacolod, Marigee P. 2007. "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality." *Review of Economics and Statistics* 89, no. 4 (November): 737-751.

Bertrand, Marianne, and Emir Kamenica. 2018. "Coming Apart? Cultural Distances in the United States Over Time." Becker Friedman Institute for Research in Economics Working Paper No. 2018-47. Chicago: (July 5).

Blagg, Kristin, and Matthew M. Chingos. 2016. *Varsity Blues: Are High School Students Being Left Behind?* Washington, DC: Urban Institute (May).

Burdick-Will, Julia, Jens Ludwig, Stephen W. Raudenbush, Robert J. Sampson, Lisa Sanbonmatsu, and Patrick Sharkey. 2011. "Converging Evidence for Neighborhood Effects on Children’s Test Scores: An Experimental, Quasi-Experimental, and Observational Comparison." In *Whither Opportunity?*, edited by Greg J. Duncan and Richard J. Murnane: Russell Sage Foundation: 255-276.

Burtless, Gary, ed. 1996. *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings.

Camera, Lauren. 2018. NAEP Shows Little to No Gains in Math, Reading for U.S. Students: Latest NAEP scores show growing gap between those at the top and bottom of the achievement spectrum. *U.S. News and World Report*, April 10.

Carneiro, Pedro, and James J. Heckman. 2003. "Human capital policy." In *Inequality in America: What role for human capital policies?*, edited by Benjamin M. Friedman. Cambridge, MA: MIT Press: 77-239.

Cheng, Albert, and Paul E. Peterson. 2018. "Experimental Estimates of Impacts of Cost-Earnings Information on Adult Aspirations for Children’s Postsecondary Education." *Journal of Higher Education*: forthcoming.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2013. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." Mimeo. Cambridge, MA: Harvard University.

Chingos, Matthew M. 2016. *No more free lunch for education policymakers and researchers*. Evidence Speaks Reports. Washington, DC: Economic Studies at Brookings (June 30).

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. 2004a. "Changing labor-market opportunities for women and the quality of teachers, 1957-2000." *American Economic Review* 94, no. 2 (May): 230-235.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. 2004b. "Women, the labor market, and the declining relative quality of teachers." *Journal of Policy Analysis and Management* 23, no. 3 (Summer): 449-470.

Cross, Christopher T. 2014. *Political education: Setting the course for state and federal policy*. Second ed. New York, NY: Teachers College, Columbia University.

Dahl, Gordon B., and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review* 102, no. 5 (August): 1927-1956.

Duncan, Greg J., Ariel Kalil, and Kathleen M. Ziol-Guest. 2017. "Increasing Inequality in Parent Incomes and Children’s Schooling." *Demography* 54: 1603-1626.

Duncan, Greg J., Pamela A. Morris, and Chris Rodrigues. 2011. "Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments." *Developmental Psychology* 47, no. 5 (September): 1263-1279.

Duncan, Greg J., and Richard J. Murnane. 2014. *Restoring opportunity: The crisis of inequality and the challenge for American education*. Cambridge, MA: Harvard Education Press.

Duncan, Greg J., and Richard J. Murnane, eds. 2011. *Whither opportunity? Rising inequality, schools, and children's life chances*. New York: Russell Sage Foundation.

Edsell, Thomas. 2012. "The Reproduction of Privilege." *New York Times*: March 12.

Egalite, Ann J. 2016. "How Family Background Influences Student Achievement." *Education Next* 16, no. 2 (Spring): 70-78.

Eide, Eric, Dan Goldhaber, and Dominic Brewer. 2004. "The teacher labour market and teacher quality." *Oxford Review of Economic Policy* 20, no. 2: 230-244.

Fetters, Williamn B., Peter S. Stowe, and Jeffrey A. Owings. 1984. *Quality of Responses of High School Students to Questionnaire Items*. High School and Beyond: a national longitudinal study for the 1980's. Washington, DC: National Center for Education Statistics (September).

Figlio, David, and Susanna Loeb. 2011. "School accountability." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 383-421.

Friedman-Krauss, Allison H., W. Steven Barnett, G.G. Weisenfeld, Richard Kasmin, Nicole DiCrecchio, and Michelle Horowit. 2018. *The state of preschool 2017: State preschool yearbook*: Rutgers University.

Friedman, Benjamin M., ed. 2003. *Inequality in America: What role for human capital policies?* Cambridge, MA: MIT Press.

Ganzeboom, Harry B. G., and Donald J. Treiman. 2003. "Three internationally standardised measures for comparative research on occupational status." In *Advances in Cross-National Comparison, A European Working Book for Demographic and Socio-Economic Variables*, edited by J.H.P. Hoffmeyer-Zlotnik and C. Wolf. New York: Kluwer Academic Press: 159-193.

Gould, Eric, Avi Simhon, and Bruce A. Weinberg. 2019. "Does Parental Quality Matter? Evidence on the Transmission of Human Capital Using Variation in Parental Influence from Death, Divorce, and

Family Size." NBER Working Paper No. 25495. Cambridge, MA: National Bureau of Economic Research (January).

Goyette, Kimberly A. 2008. "College for some to college for all: Social background, occupational expectations, and educational expectations over time." *Social Science Research* 37, no. 2 (June): 461-484.

Greenberg, Erica. 2018. *New Measures of Student Poverty: Replacing Free and Reduced-Price Lunch Status Based on Household Forms with Direct Certification* Wasnington, DC: Urban Institute (November).

Grissmer, David, Ann Flanagan, and Stephanie Williamson. 1998. "Why did the Black-white score gap narrow in the 1970s and 1980s?" In *The Black-white test score gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution Press: 181-226.

Grissmer, David W., Sheila Nataraj Kirby, Mark Berends, and Stephanie Williamson. 1994. *Student achievement and the changing American family*. Santa Monica, CA: Rand Corporation.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Symposia: Investment in Children: Parental Education and Parental Time with Children." *Journal of Economic Perspectives* 22, no. 3 (Summer): 23-46.

Hanushek, Eric A. 1998. "Conclusions and controversies about the effectiveness of school resources." *FRBNY Economic Policy Review* 4(March): 11-28.

Hanushek, Eric A. 2016a. "School human capital and teacher salary policies." *Journal of Professional Capital and Community* 1, no. 1: 23-40.

Hanushek, Eric A. 2016b. "What Matters for Achievement: Updating Coleman on the Influence of Families and Schools." *Education Next* 16, no. 2 (Spring): 22-30.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39, no. 2 (Spring): 326-354.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2009. "New evidence about *Brown v. Board of Education*: The complex effects of school racial composition on achievement." *Journal of Labor Economics* 27, no. 3 (July): 349-383.

Hanushek, Eric A., and Alfred A. Lindseth. 2009. *Schoolhouses, courthouses, and statehouses: Solving the funding-achievement puzzle in America's public schools*. Princeton, NJ: Princeton University Press.

Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann. 2012. *Achievement growth: International and U.S. state trends in student achievement*. Cambridge, MA: Program on Education Policy and Governance, Harvard Kennedy School.

Hanushek, Eric A., Marc Piopiunik, and Simon Wiederhold. forthcoming. "The value of smarter teachers: International evidence on teacher cognitive skills and student performance." *Journal of Human Resources*.

Hanushek, Eric A., and Margaret E. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management* 24, no. 2: 297-327.

Hanushek, Eric A., and Steven G. Rivkin. 1997. "Understanding the twentieth-century growth in U.S. school spending." *Journal of Human Resources* 32, no. 1 (Winter): 35-68.

Hanushek, Eric A., and Steven G. Rivkin. 2006. "Teacher quality." In *Handbook of the Economics of Education, Vol. 2*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 1051-1078.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): 267-271.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2015. "Returns to skills around the world: Evidence from PIAAC." *European Economic Review* 73: 103-130.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2017. "Coping with change: International differences in the returns to skills." *Economic Letters* 153(April): 15-19.

Hanushek, Eric A., and Ludger Woessmann. 2011. "The economics of international differences in educational achievement." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 89-200.

Hart, Betty, and Todd R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children* Baltimore, MD: Paul H. Brookes Publishing Co.

Hart, Betty, and Todd R. Risley. 2003. "The early catastrophe: The 20 million word gap by age 3." *American Educator*(Spring).

Hedges, Larry V., and Amy Nowell. 1998. "Black-white test score convergence since 1965." In *The Black-white test score gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution Presss: 149-181.

Hoff, Erika. 2003. "The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development via Maternal Speech." *Child Development* 74, no. 5 (Sepember-October): 1368-1378.

Hoxby, Caroline M. 2003. "School choice and school competition: Evidence from the United States." *Swedish Economic Policy Review* 10, no. 3: 9-65.

Hoxby, Caroline M., and Andrew Leigh. 2004. "Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States." *American Economic Review* 94, no. 2 (May): 236-240.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *Quarterly Journal of Economics* 131, no. 1 (February): 157-218.

Jacob, Brian A., and Tamara Wilder Linkow. 2011. "Educational Expectations and Attainment." In *Whither Opportunity?*, edited by Greg J. Duncan and Richard J. Murnane: Russell Sage Foundation: 133-162.

Jencks, Christopher, and Meredith Phillips, eds. 1998. *The black-white test score gap*. Washington, D.C.: The Brookings Institution.

Kalogrides, Demetra, Susanna Loeb, and Tara Béteille. 2013. "Systematic Sorting: Teacher Characteristics and Class Assignments." *Sociology of Education* 86, no. 2 (April): 103-123.

Kao, Grace, and Marta Tienda. 1998. "Educational Aspirations of Minority Youth." *American Journal of Education* 106, no. 3 (June): 349-384.

Kaufman, Philip, and Kenneth A. Rasinski. 1991. *Quality of the Responses of Eighth-Grade Students in NELS:88*. National Education Longitudinal Study of 1988. Washington, DC: National Center for Education Statistics (September).

Kayser, Brian D., and Gene F. Summers. 1973. "The Adequacy of Student Reports of Parental Ses Characteristics." *Sociological Methods & Research* 1, no. 3: 303-315.

Krueger, Alan B. 1998. "Reassessing the view that American schools are broken." *FRBNY Economic Policy Review* 4, no. 1 (March): 29-43.

Krueger, Alan B. 2003. "Inequality, too much of a good thing." In *Inequality in America: What role for human capital policies?*, edited by Benjamin M. Friedman. Cambridge, MA: MIT Press: 1-75.

Ladd, Helen F., ed. 1996. *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D.C.: The Brookings Institution.

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

Loeb, Susanna, Demetra Kalogrides, and Tara Béteille. 2012. "Effective Schools: Teacher Hiring, Assignment, Development, and Retention." *Education Finance and Policy* 7, no. 3 (Summer): 269–304.

Magnuson, Katherine, and Jane Waldfogel, eds. 2008. *Steady gains and stalled progress: Inequality and the black-white test score gap*. New York: Russell Sage Foundation.

Maxie, Tre. 2012. "Education Challenge for the Next 50 Years is the Achievement Gap." *Seattle Times*: September 28.

Mayer, Susan E. 1997. *What money can't buy: Family income and children's life chances*. Cambridge, MA: Harvard University Press.

Morgan, Paul L., George Farkas, Marianne M. Hillemeier, and Steve Maczuga. 2017. "Replicated Evidence of Racial and Ethnic Disparities in Disability Identification in U.S. Schools." *Educational Researcher* 46, no. 6 (August/September): 305-322.

Murray, Charles. 2012. *Coming apart: The state of white America, 1960-2010*. New York: Crown Forum.

National Center for Education Statistics. 2012. *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*. Washington, D.C.

National Center for Education Statistics. 2013. *The Nation's Report Card: Trends in Academic Progress 2012*. Washington, DC: Institute of Education Sciences.

Neff, Walter S. 1938. "Socioeconomic status and intelligence: A critical survey." *Psychological Bulletin* 35, no. 10 (December): 727-757.

Nelson, Charles A., and Margaret A. Sheridan. 2011. "Lessons from Neuroscience Research for Understanding Causal Links Between Family and Neighborhood Characteristics and Educational Outcomes." In *Whither Opportunity?*, edited by Greg J. Duncan and Richard J. Murnane: Russell Sage Foundation: 27-46.

Nielsen, Eric R. 2015. "The Income-Achievement Gap and Adult Outcome Inequality." Finance and Economics Discussion Series 2015-041. Washington, DC: Federal Reserve Board.

OECD. 2017a. *PISA 2015 results (volume III): Students' well-being*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2017b. *PISA 2015 Technical Report*. Paris: Organisation for Economic Co-operation and Development.

Perna, Laura W. 2006. "Studying College Access and Choice: A Proposed Conceptual Model." In *Higher Education: Handbook of Theory and Research*, edited by John C. Smart. Dordrecht: Springer Netherlands: 99-157.

Peterson, Paul E. 2010. *Saving schools: From Horace Mann to virtual learning*. Cambridge, MA: Belknap Press of Harvard University Press.

Peterson, Paul E., and Martin R. West, eds. 2007. *School money trials*. Washington: Brookings.

Putnam, Robert D. 2015. *Our kids: The American dream in crisis*. New York: Simon & Schuster.

Reardon, Sean F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, edited by Richard J. Murnane and Greg J. Duncan. New York: Russell Sage Foundation Press: 91-116.

Rivkin, Steve G. 2016. "Desegregation since the Coleman Report: Racial composition of schools and student learning." *Education Next* 16, no. 2 (Spring): 29-37.

Rivkin, Steven G., and Finis Welch. 2006. "Has school desegregation improved academic and economic outcomes for blacks?" In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 1019-1049.

Rothstein, Richard. 2004. *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. Washington, DC: Economic Policy Institute.

Saez, Emmanuel, and Gabriel Zucman. 2016. "wealth inequality in the United States since 1913: Evidence from capitalized income tax data." *Quarterly Journal of Economics* 131, no. 2 (May): 519-578.

Schofield, Janet Ward. 1995. "Review of research on school desegregation's impact on elementary and secondary school students." In *Handbook of research on multicultural education*, edited by James A. Banks and Cherry A. McGee Banks. New York: Macmillan Publishing: 597-616.

Slesnick, Daniel T. 1993. "Gaining Ground: Poverty in the Postwar United States." *Journal of Political Economy* 101, no. 1 (February): 1-38.

Smith, Marshall. 1972. "Equality of educational opportunity: the basic findings reconsidered." In *On equality of educational opportunity*, edited by Frederick Mosteller and Daniel P. Moynihan. New York: Random House: 230-342.

Taverise, Sabrina. 2012. "Education Gap Grows Between Rich and Poor, Studies Say." *New York Times*: February 9.

U.S. Department of Education. 2013. *Digest of Education Statistics, 2012*. Washington, DC: National Center for Education Statistics.

Weissmann, Jordan. 2012. "Occupy Kindergarten." *Atlantic*: Feb. 11.

Welch, Finis, and Audrey Light. 1987. *New evidence on school desegregation*. Washington, D.C.: U. S. Commission on Civil Rights.

Wilson, William Julius. 1987. *The truly disadvantaged: The innercity, the underclass, and public policy*. Chicago: University of Chicago Press.

Wilson, William Julius. 2011. ""The Declining Significance of Race": Revisited & Revised." *Daedalus* 140, no. 2 (Spring): 55-69.

Wilson, William Julius. 2012. *The declining significance of race: Blacks and changing American institutions*. 3rd ed. Chicago: University of Chicago Press.

Zernicke, Kate. 2016. "Test Scores Show a Decline in Math Among High School Seniors." *New York Times*: April 27.

Appendix A: Data Sources for Educational Achievement

We use four surveys to investigate achievement gaps over time: two assessments of the National Assessment of Educational Progress (NAEP), the Trends in International Mathematics and Science Study (TIMSS), and the Programme for International Student Assessment (PISA). All four are widely used in studying long-term trends in student achievement. Appendix Table A.1 indicates the specific years in which the different surveys were administered. Appendix Table A.2 provides the data used in our trend analyses.

Each dataset is comprised of microdata at the student level, which we aggregate by demographic groups. PISA and TIMSS national microdata are available to the public on the website of the National Center for Education Statistics (NCES), but to use NAEP microdata the user must gain access to restricted-use data files.

All four exams include student questionnaires that include questions about students' background, attitudes, and experiences in school. Questionnaire responses are linked to students' test scores for each subject. We combine these data to study achievement trends by groups of students.

LTT-NAEP

We use two datasets provided by NAEP and treat them separately. The Long-Term Trend (LTT) assessment dates back to 1969 and assesses students aged 9, 13, and 17 years. LTT-NAEP data are available for math in select years from 1978-2008 and for reading from 1971-2008. We create a panel of math and reading scores for 8^{th} and 12^{th} graders.

Main-NAEP

Main-NAEP assesses students in grades 4, 8, and 12. Main-NAEP trend data are available for select years in 1990-2013; we create a panel of math and reading scores for 8^{th} graders. All NAEP data come from the National Center for Education Statistics (NCES) and were analyzed in a restricted-use data room.

TIMSS

TIMSS assesses 4^{th} and 8^{th} graders in math and science, and there are data available every four years from 1995-2015. We create a panel of 8^{th} grade microdata using national data files from 2003, 2007, and 2011, and international data files from 1995, 1999, and 2015. The only apparent difference between our national and international data years is that the international data do not contain an indicator of race or ethnicity. For this reason, our estimates of the achievement gap by race for TIMSS are only available for 2003, 2007, and 2011.

PISA

Rather than testing children at certain grade levels, PISA assesses math and reading in children at age 15. By testing children who are nearing the end of their compulsory schooling in most countries, it attempts to measure the "yield" of a country's education system. We use national PISA data, available every three years from 2000-2015.

Appendix B: Measuring Socio-economic Status

To be able to observe percentiles of the SES distribution in each survey and year, we construct a continuous measure of SES. A single composite measure of SES allows us to identify the inter-quartile range of the SES distribution and identify the 90-10 SES-achievement gap, which provides a clearer picture of the impact of SES on student achievement than the use of ever-changing categorical groups. None of the intertemporally linked surveys include indicators of earned income or other household receipts other than the free and reduced lunch indicators in NAEP surveys, and only the PISA survey contains information on parental occupation. Thus, we measure SES by use of an index that includes levels of parental educational attainment and the amount and variety of durable and educational goods available within the household. In a separate survey with parent-reported income data, the index is highly correlated with an estimate of permanent income.

B.1 The PISA Index of Economic, Social, and Cultural Status (ESCS)

Across the different PISA waves, the OECD provides a measure of socio-economic status called the PISA Index of Economic, Social, and Cultural Status (ESCS). The ESCS, according to the PISA 2015 Technical Report, is "a composite score built by the indicators parental education (*pared*), prestige of the occupation of the parent with the highest occupational ranking (*hisei*), and home possessions (*homepos*) including books in the home via principal component analysis (PCA)…. The rationale for using these three components was that socio-economic status has usually been seen as based on education, occupational status and income. As no direct income measure has been available from the PISA data, the existence of household items has been used as a proxy for family wealth" (OECD (2017b)).

To compute the ESCS index, PISA uses a combination of highest parental education (in years), parental occupation (transformed into an International Socio-Economic Index of Occupational Status (ISEI), see Ganzeboom and Treiman (2003)), and home possessions (derived from 10-15 yes/no questions such as "do you have a computer in your home?" and 3-5 questions such as "how many cars does your family own?"). PISA standardizes the three variables, performs Principal Components Analysis (PCA), and defines ESCS as the component score for the first principal component. Materials in the home included in PISA 2000 included the following: dishwasher, own bedroom, educational software, a link to the Internet, a dictionary, a quiet place to study, a desk, textbooks, classic literature, books of poetry, and works of art. In PISA 2015, the items included the number of personal computers and cell phones in the home.

The benefit of using this method to investigate trends by socio-economic status, rather than simply using one or a combination of categorical variables like eligibility for national school lunch programs, parent education, or books in home, is that it can account for changes in the share of students within these categories over time. In any of these categorical variables, shifts in culture and technology can alter the distribution of students between categories over time, reducing the validity of their use as proxies for SES. For example, the proportion of students having no books in their home versus over 200 books in their home has changed dramatically during the past fifteen years (Appendix Figure B.1, Panel A). Meanwhile, the proportion of children with internet access has increased to almost 100 percent, rendering the variable useless if used on its own (Appendix Figure B.1, Panel B).

B.2 Our SES Index

In the construction of our SES index, we follow closely the spirit of PISA's ESCS index, making appropriate adjustments to enable implementation in all our four surveys. Neither NAEP nor TIMSS provide a similar index, although NAEP is considering adding a similar measure to their series (see National Center for Education Statistics (2012)). Therefore, we construct a comparable SES index for the four underlying surveys ourselves. While we have to make adjustments because the other surveys in our analysis do not include all the information available to PISA, using the PISA data we show that our SES index is highly correlated with PISA's ESCS index.

Our SES index differs from the PISA ESCS index in the following ways.

Parental education. Instead of using the highest parental education in years (*pared*), we use the categorical variable of highest parental education (*hisced*) to construct our index. *Hisced* and *pared* have the exact same distribution, but instead of being measured in years of education, *hisced* is measured categorically on the International Standard Classification of Education (ISCED). We choose to use *hisced* instead of *pared* for consistency with the other two assessments, which both measure highest parental education on the ISCED scale, so that we do not have to rely on a potentially error-prone transformation into years of education.

Parental occupation. Unlike PISA, the student questionnaires in NAEP and TIMSS do not include measures of the parents' occupations that would allow for estimating occupational prestige (*hisei*). We therefore exclude measures of parents' occupations from our index. Though it is unfortunate to lose this element in our measure of socio-economic status, the category is largely redundant of the education and income items that remain in the index, as the prestige of an occupation is estimated from the education and income of the average member of the occupation. Estimations of the SES-achievement gap in the PISA data set closely resemble estimates obtained when PISA's ESCS index is employed (see below).

Home possessions. To create ESCS, the OECD uses an index of home possessions (*homepos*) which is "a summary index of all household and possessions items" (OECD (2017b)). NAEP and TIMSS include similar questions about students' home possessions, but they do not provide a summary index.[23] For all estimations of SES, we therefore use a simple sum of the home possessions variables as our indicator of home possessions (*homepos*). That is, we simply add up each of the home possessions students reported owning and used this number as our *homepos* variable in the specific survey and year.[24]

Construction of the index. Using *homepos* and *hisced*, we simply follow the ESCS construction process of performing PCA and assigning each student a composite score.

[23] NAEP surveys also ask different questions about home possessions across years. Generally speaking, they ask ten to fifteen yes/no questions such as "do you have a computer in your home?" and three to five questions where answers can vary across a continuum, such as "how many cars does your family own?"

[24] Because some home possessions variables are missing for some students, we also considered computing *homepos* as a ratio of owned items to known items. In this case, *homepos* would be the sum of items possessed divided by the number of non-missing items. We did not make this adjustment, as it had a slightly lower correlation with the ESCS index.

In the construction, we differ slightly from the ESCS process in the treatment of missing variables. The OECD treats missing variables in the following way: "For students with missing data on one out of the three components, the missing variable was imputed. Regression on the other two variables was used to predict the third (missing) variable, and a random component was added to the predicted value. If there were missing data on more than one component, ESCS was not computed and a missing value was assigned for ESCS" (OECD (2017b)). As this method requires the assumption of a positive, linear relationship among the variables and in any case only applied to 2% of the observations, instead of imputing missing variables we choose to discard them from the analysis.

Comparing our SES index to the PISA ESCS index. The joint impact of these alterations is the construction of an index that remains highly correlated with the PISA ESCS index. When we calculate both our SES index and PISA's ESCS index within the same PISA data set, the overall correlation between the two is 0.876. It ranges from 0.87 to 0.91 when broken down by years.

Because we are interested in examining trends for students at the extremes of the distribution, we compared trends in the 90-10 gap in PISA using both the ESCS and our SES index. No qualitatively significant differences between the trends estimated by the two indices are observed (see Appendix Figure B.2).

B.3 SES Index and Earned Income

To estimate the relationship between our index and family income, we use data from the 1988 and 2002 Education Longitudinal Study (ELS), which contain home possessions variables (quite similar to those in PISA), parent education, and income. Annual income, obtained from parent questionnaires, is defined as "total family income from all sources [for the previous calendar year]", reported in thirteen categories ranging from "None" to "$200,001 or more." In the 1988 ELS, family income is available on the base year survey (1987 income) and on the second follow-up survey (1991 income). We built the SES index in the same way as in our main analysis (using home possessions and parent education).

The correlation between the SES index and reported family income is displayed in Appendix Table B.1. The two variables are strongly but not perfectly correlated. Interestingly enough, at 0.66 the SES index is more highly correlated with the average of the annual earnings estimates obtained in 1987 and 1991 than with either of the annual estimates, suggesting that the average is a better measure of permanent income, a concept similar to socio-economic status.[25]

[25] Using 2002 ELS data, where family income is available only on the base year survey (2001 income), the correlation between the SES index and reported income is 0.503.

Table A.1: Survey and Subject by Test Date, 1971-2015

	LTT-NAEP				Main-NAEP		PISA			TIMSS	
	13-year-olds		17-year-olds		8th graders		15-year-olds			8th graders	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Science	Math	Science
1971		X		X							
1973	X		X								
1975		X		X							
1978	X		X								
1980		X		X							
1982	X		X								
1986	X		X								
1988		X		X							
1990	X	X	X	X	X	X					
1991											
1992	X	X	X	X	X	X					
1993											
1994	X	X	X	X		X					
1995										X	X
1996	X	X	X	X	X						
1997											
1998						X					
1999	X	X	X	X						X	X
2000					X		X	X	X		
2001											
2002						X					
2003							X	X	X	X	X
2004	X	X	X	X							
2005					X	X					
2006							X		X		
2007					X	X				X	X
2008	X	X	X	X							
2009					X	X	X	X	X		
2010											
2011					X	X				X	X
2012	X	X	X	X			X	X	X		
2013					X	X					
2014											
2015					X	X	X	X	X	X	X

Note: LTT-NAEP math data for 1973 are available for levels but not gaps.

Table A.2: Data for Trend Analyses by Survey, Test Year, Age, and Subject

					Unconditional gaps		SES gaps	
Test	Test year	Age	Subject	Mean	90-10	75-25	90-10	75-25
pisa	2000	15	math	493	2.622	1.364	1.572	1.273
pisa	2003	15	math	482	2.564	1.329	1.325	1.050
pisa	2006	15	math	474	2.394	1.293	1.249	0.946
pisa	2009	15	math	488	2.429	1.302	1.185	0.936
pisa	2012	15	math	481	2.371	1.285	1.140	0.885
pisa	2015	15	math	470	2.386	1.262	0.999	0.802
pisa	2000	15	reading	504	2.632	1.372	1.469	1.133
pisa	2003	15	reading	494	2.495	1.332	1.287	1.012
pisa	2009	15	reading	500	2.423	1.304	1.188	0.932
pisa	2012	15	reading	497	2.237	1.214	1.016	0.793
pisa	2015	15	reading	498	2.497	1.347	0.806	0.649
pisa	2000	15	science	499	2.525	1.409	1.487	1.153
pisa	2003	15	science	490	2.560	1.410	1.308	1.041
pisa	2006	15	science	489	2.733	1.522	1.374	1.058
pisa	2009	15	science	502	2.487	1.355	1.255	0.955
pisa	2012	15	science	497	2.349	1.298	1.154	0.868
pisa	2015	15	science	496	2.502	1.384	0.923	0.753
timss	1995	14	math	487	2.577	1.413	0.741	0.653
timss	1999	14	math	502	2.463	1.327	0.847	0.685
timss	2003	14	math	504	2.314	1.222	0.915	0.677
timss	2007	14	math	509	2.213	1.190	0.901	0.616
timss	2011	14	math	509	2.199	1.154	0.897	0.620
timss	2015	14	math	518	2.406	1.286	0.945	0.718
timss	1995	14	science	521	2.588	1.365	0.642	0.563
timss	1999	14	science	515	2.379	1.257	0.916	0.684
timss	2003	14	science	527	1.991	1.048	0.844	0.620
timss	2007	14	science	520	2.002	1.074	0.939	0.608
timss	2011	14	science	524	1.982	1.053	0.914	0.589
timss	2015	14	science	529	2.007	1.059	0.818	0.616
naep	1990	14	math	259	2.589	1.427	1.198	1.045
naep	1992	14	math	262	2.940	1.590	1.255	1.056
naep	1996	14	math	271	2.785	1.487	1.126	0.943
naep	2000	14	math	276	2.817	1.475	1.201	1.018
naep	2005	14	math	279	2.774	1.454	1.450	1.105
naep	2007	14	math	281	2.742	1.443	1.455	1.092
naep	2009	14	math	283	2.776	1.456	1.315	1.163
naep	2011	14	math	284	2.767	1.458	1.458	1.121
naep	2013	14	math	285	2.771	1.460	1.436	1.037
naep	2015	14	math	282	2.832	1.496	1.467	1.077
naep	1990	14	reading	255	2.554	1.345	0.955	0.825
naep	1992	14	reading	254	2.542	1.337	1.050	0.852
naep	1994	14	reading	254	2.554	1.355	1.085	0.888
naep	1998	14	reading	263	2.320	1.199	1.067	0.842
naep	2002	14	reading	264	2.228	1.156	1.137	0.843
naep	2005	14	reading	262	2.351	1.219	1.202	0.907
naep	2007	14	reading	263	2.303	1.177	1.212	0.892
naep	2009	14	reading	264	2.267	1.160	1.079	0.923
naep	2011	14	reading	265	2.269	1.172	1.235	0.919
naep	2013	14	reading	268	2.280	1.180	1.230	0.842

Table A.2 (continued)

Test	Test year	Age	Subject	Mean	Unconditional gaps 90-10	Unconditional gaps 75-25	SES gaps 90-10	SES gaps 75-25
naep	2015	14	reading	265	2.321	1.188	1.201	0.847
naepltt	1978	13	math	264	2.846	1.532	1.372	0.999
naepltt	1982	13	math	268	2.442	1.298	1.040	0.729
naepltt	1986	13	math	269	2.276	1.183	0.966	0.819
naepltt	1990	13	math	270	2.314	1.208	0.933	0.805
naepltt	1992	13	math	273	2.273	1.202	1.038	0.830
naepltt	1994	13	math	275	2.372	1.236	1.065	0.880
naepltt	1996	13	math	275	2.334	1.209	1.039	0.847
naepltt	1999	13	math	275	2.415	1.256	1.082	0.923
naepltt	2004	13	math	281	2.399	1.247	1.127	0.872
naepltt	2008	13	math	281	2.470	1.253	1.061	0.847
naepltt	2012	13	math	285	2.535	1.318	1.125	0.934
naepltt	1971	13	reading	255	2.025	1.053	0.959	0.740
naepltt	1975	13	reading	256	2.026	1.027	0.955	0.719
naepltt	1980	13	reading	258	1.944	1.041	0.843	0.685
naepltt	1988	13	reading	258	1.907	1.025	0.610	0.479
naepltt	1990	13	reading	257	2.029	1.050	0.774	0.614
naepltt	1992	13	reading	260	2.218	1.162	0.889	0.744
naepltt	1994	13	reading	258	2.224	1.141	0.903	0.682
naepltt	1996	13	reading	258	2.228	1.135	0.891	0.737
naepltt	1999	13	reading	259	2.156	1.146	0.888	0.705
naepltt	2004	13	reading	259	2.054	1.070	0.843	0.563
naepltt	2008	13	reading	260	2.072	1.045	0.990	0.683
naepltt	2012	13	reading	263	2.067	1.062	0.868	0.721
naepltt	1973	14	math	266				
naepltt	1973	17	math	304				
naepltt	1978	17	math	300	2.580	1.408	1.251	0.993
naepltt	1982	17	math	298	2.425	1.305	1.050	0.808
naepltt	1986	17	math	302	2.305	1.251	1.100	0.921
naepltt	1990	17	math	305	2.314	1.287	1.010	0.891
naepltt	1992	17	math	307	2.215	1.196	0.937	0.746
naepltt	1994	17	math	306	2.278	1.178	1.018	0.815
naepltt	1996	17	math	307	2.264	1.203	0.943	0.690
naepltt	1999	17	math	308	2.328	1.246	0.875	0.717
naepltt	2004	17	math	307	2.172	1.152	0.964	0.852
naepltt	2008	17	math	306	2.174	1.164	0.952	0.752
naepltt	2012	17	math	306	2.246	1.169	1.027	0.840
naepltt	1971	17	reading	286	2.536	1.327	1.220	0.901
naepltt	1975	17	reading	285	2.447	1.278	1.146	0.864
naepltt	1980	17	reading	285	2.335	1.232	1.146	0.863
naepltt	1988	17	reading	290	2.110	1.115	0.802	0.633
naepltt	1990	17	reading	290	2.298	1.202	0.900	0.743
naepltt	1992	17	reading	290	2.429	1.232	0.975	0.799
naepltt	1994	17	reading	288	2.466	1.290	0.983	0.797
naepltt	1996	17	reading	287	2.339	1.220	0.901	0.750
naepltt	1999	17	reading	288	2.349	1.215	0.985	0.791
naepltt	2004	17	reading	285	2.445	1.234	1.096	0.816
naepltt	2008	17	reading	286	2.484	1.268	1.084	0.773
naepltt	2012	17	reading	287	2.372	1.229	1.058	0.844

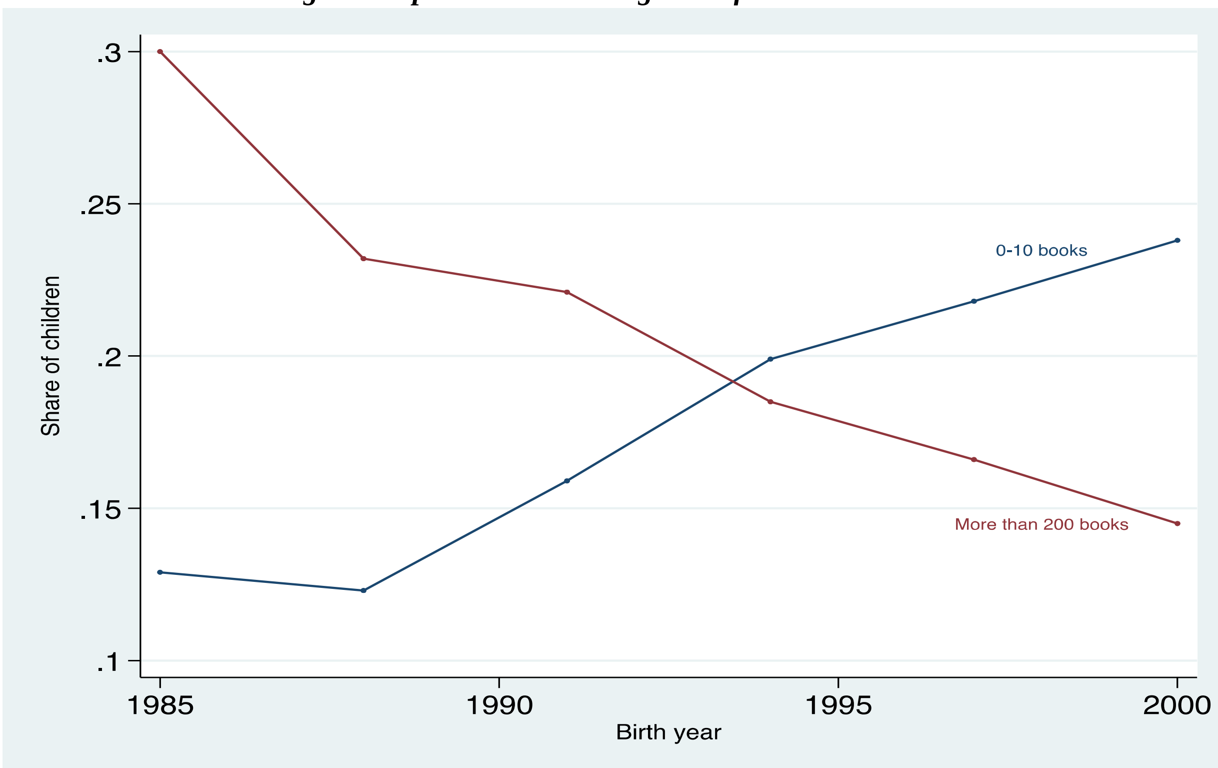
Table B.1: Correlation between SES Index and Family Income in the ELS

	SES index	1987 income	1991 income	Permanent income
SES index	1	0.51	0.59	0.66
1987 income	0.51	1	0.75	0.94
1991 income	0.59	0.75	1	0.94
Permanent income	0.66	0.94	0.94	1

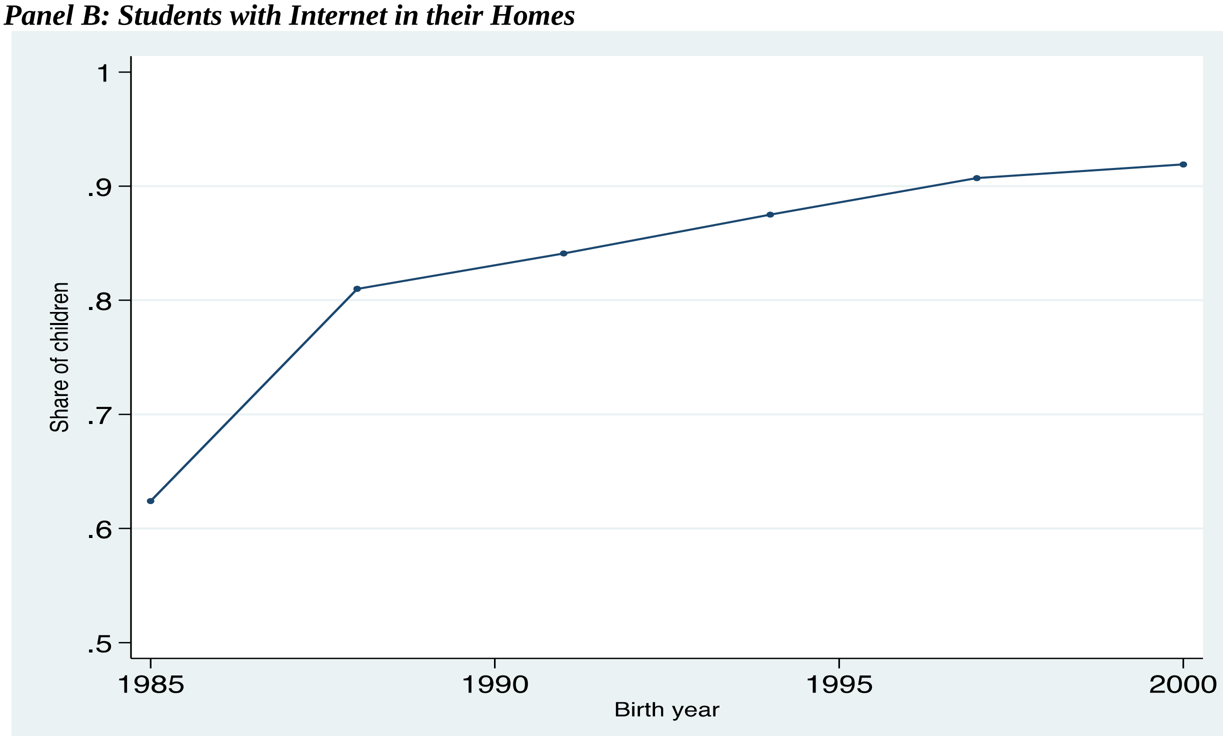
Note: Data source: 1988 Education Longitudinal Study (ELS).

Figure B.1: Changing Proportion of Students with Books and Internet in their Homes, Birth Cohorts 1985-2000

Panel A: Students Falling into Top and Bottom Categories of Books in Home

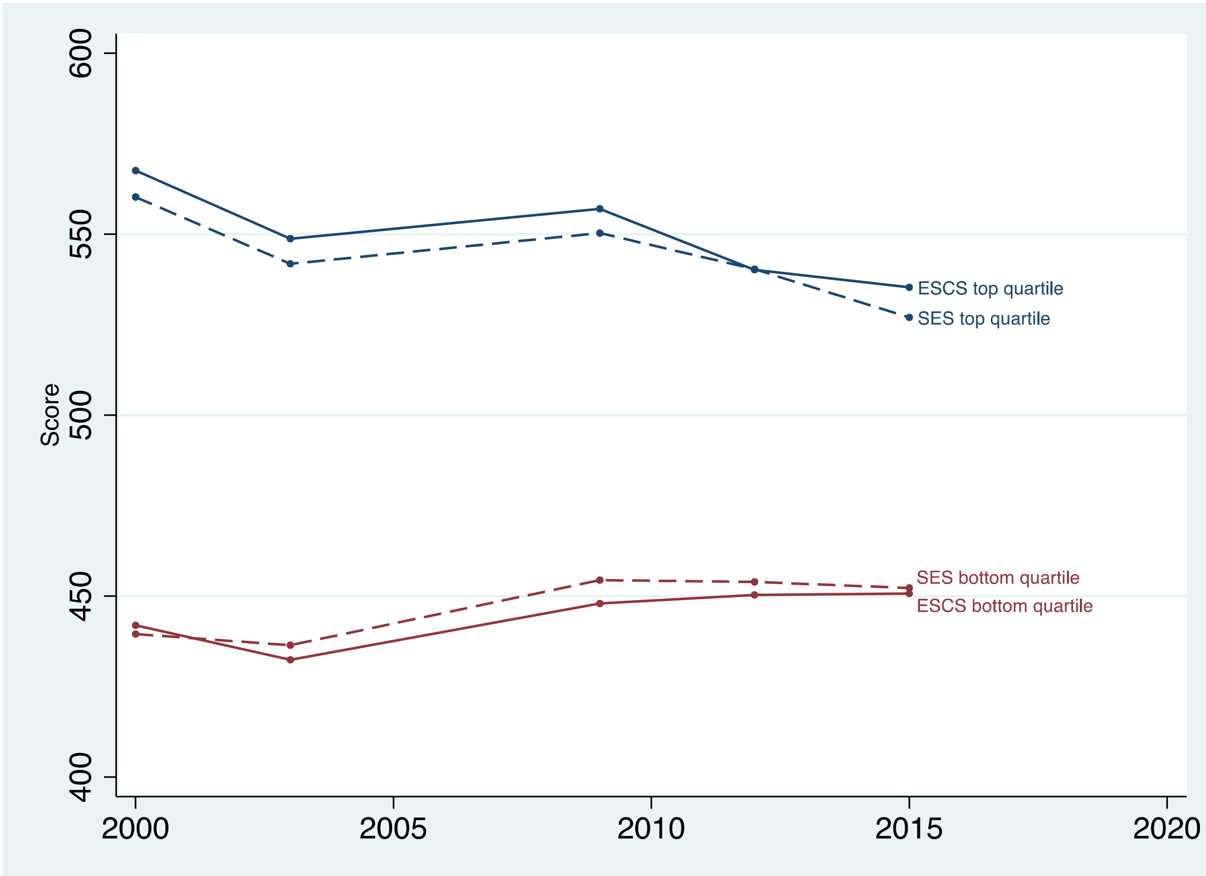


Panel B: Students with Internet in their Homes



Note: U.S. student population in PISA.

Figure B.2: Achievement Trends of the Top and Bottom Quartile in PISA based on PISA's ESCS Index and our SES Index by Test Year



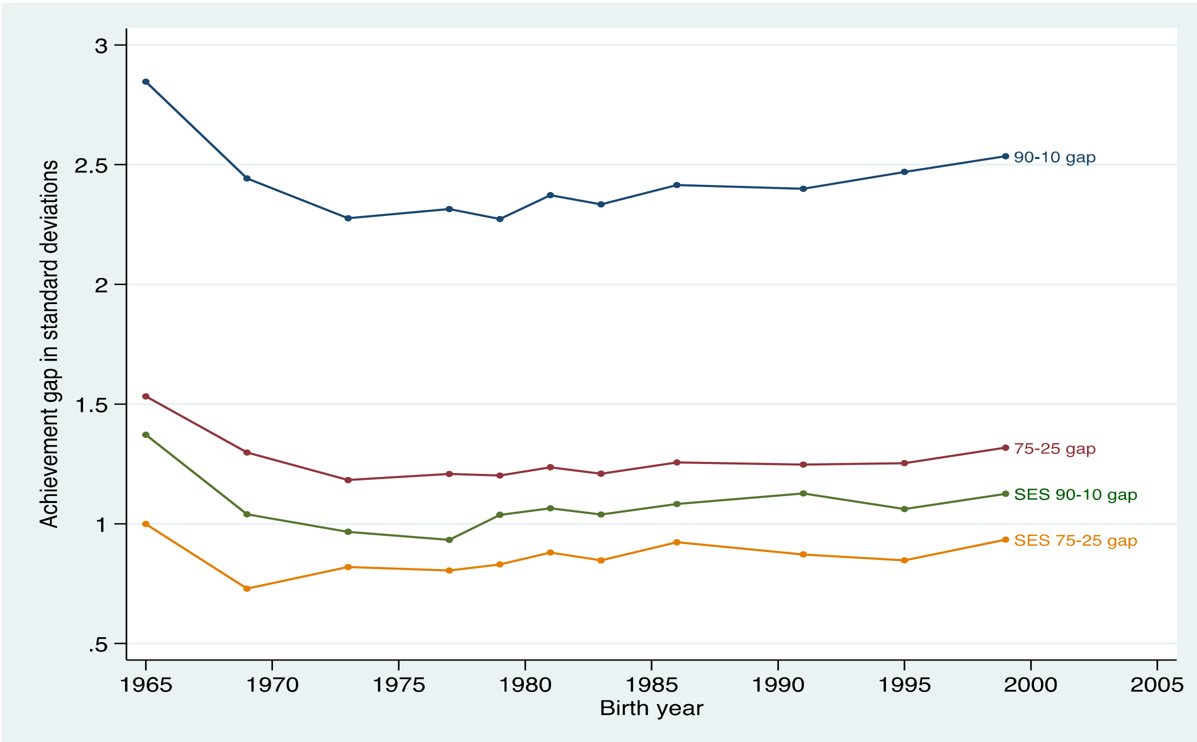
Note: U.S. student population in PISA. Each point represents roughly 400-700 students. Mean scores for the top and bottom quartiles in each index were averaged across math, reading, and science.

Appendix C: Achievement Gaps in Individual Tests

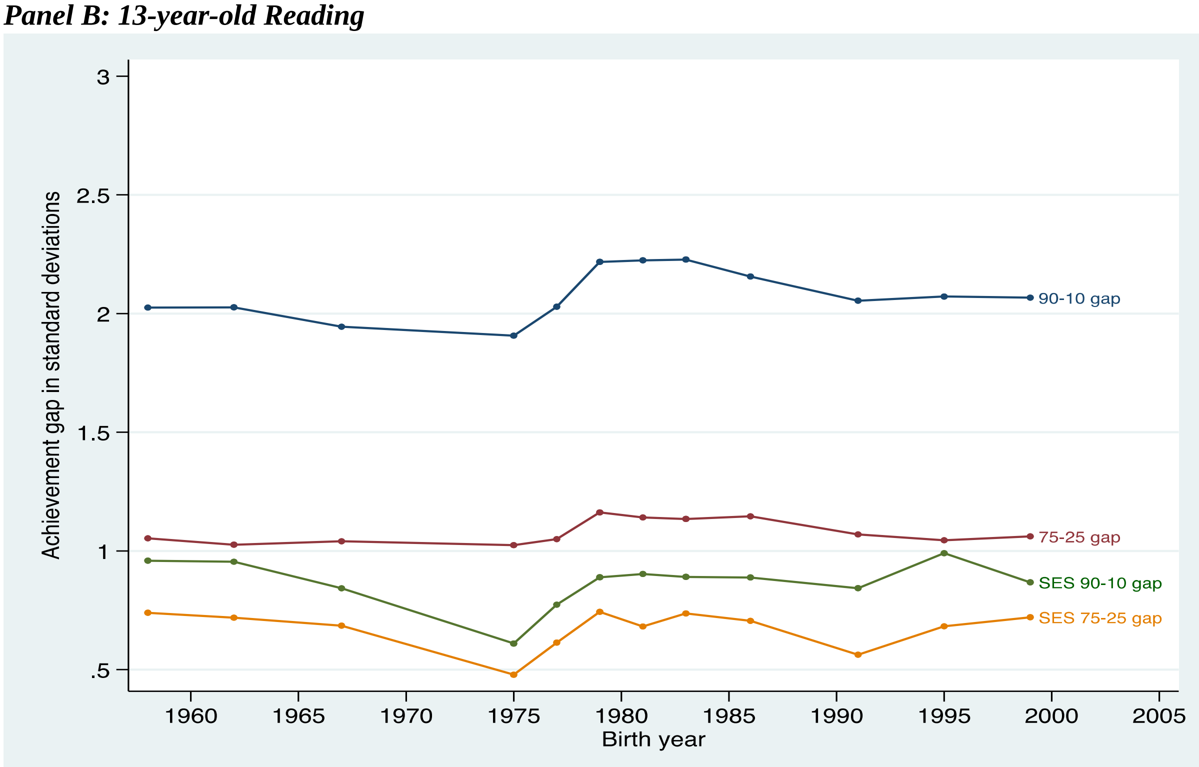
Figures C.1-C.4 plot achievement gaps in each individual test—Main-NAEP, LTT-NAEP, TIMSS, and PISA—by birth year.

Figure C.1: Unconditional and SES Achievement Gaps in LTT-NAEP, Birth Cohorts 1954-1999

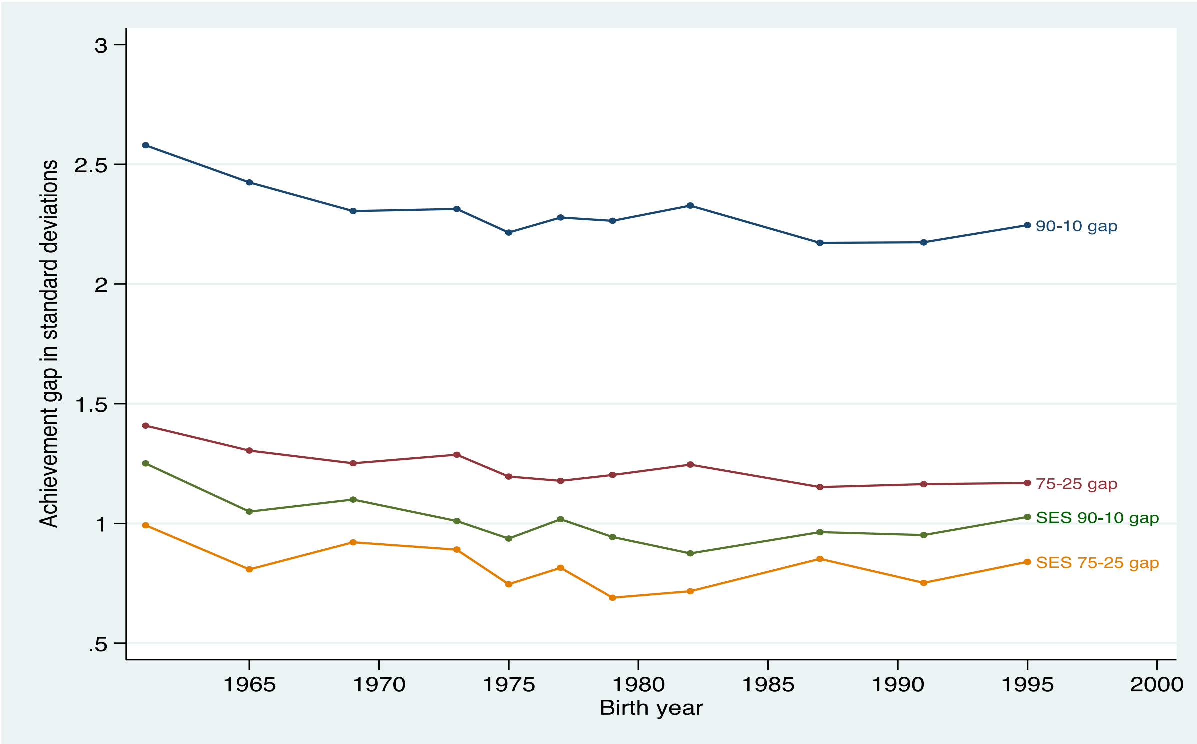
Panel A: 13-year-old Math



Panel B: 13-year-old Reading



Panel C: 17-year-old Math



Panel D: 17-year-old Reading

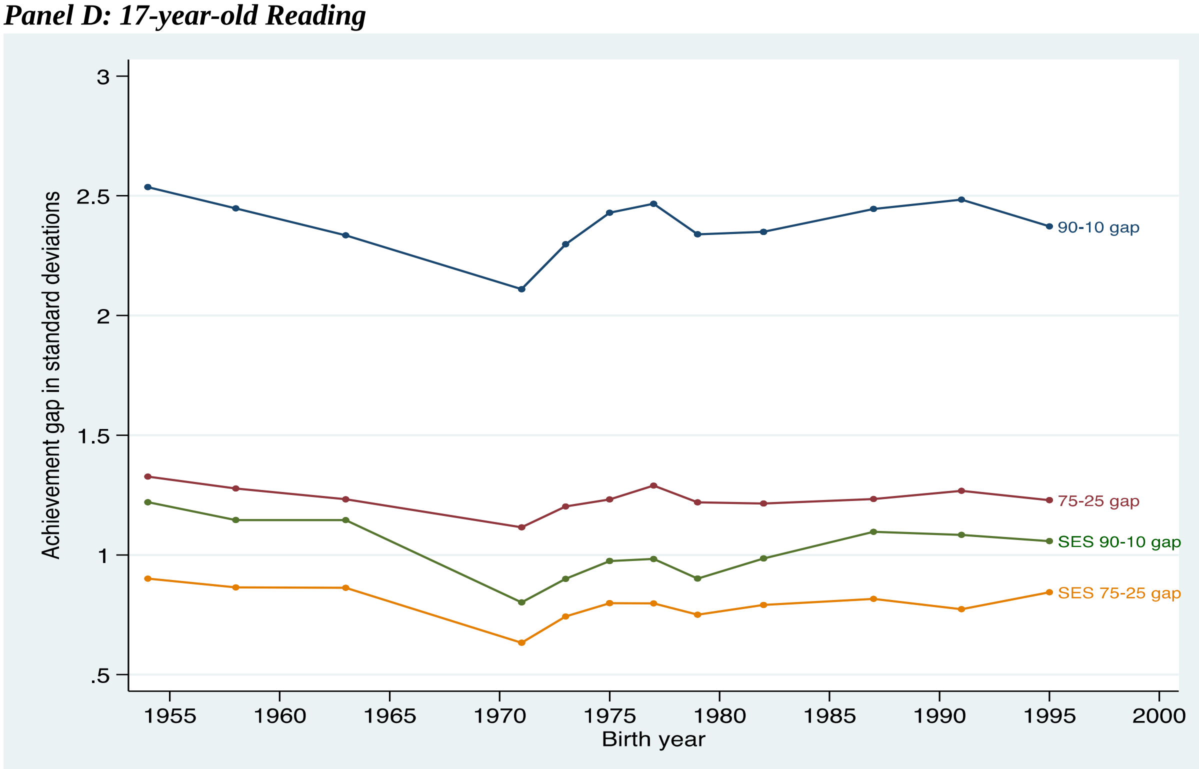
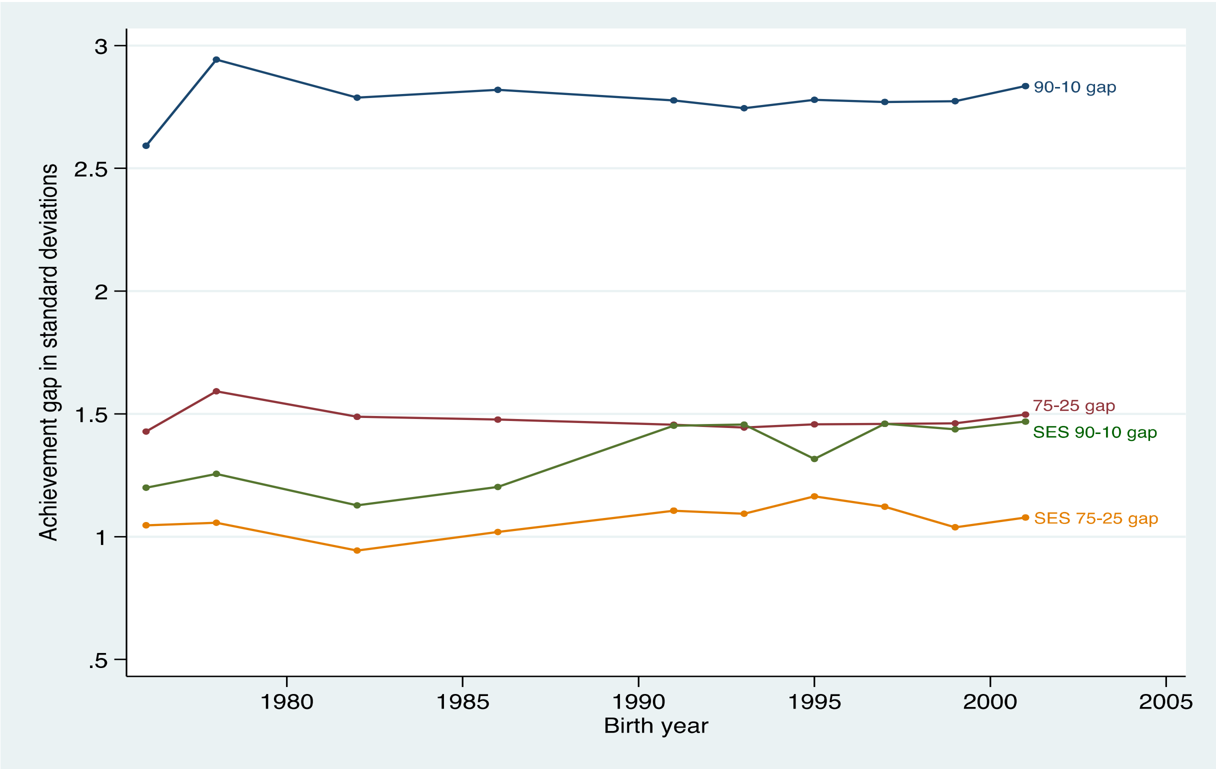


Figure C.2: Unconditional and SES Achievement Gaps in Main-NAEP, Birth Cohorts 1977-2001

Panel A: 8th Grade Math



Panel B: 8th Grade Reading

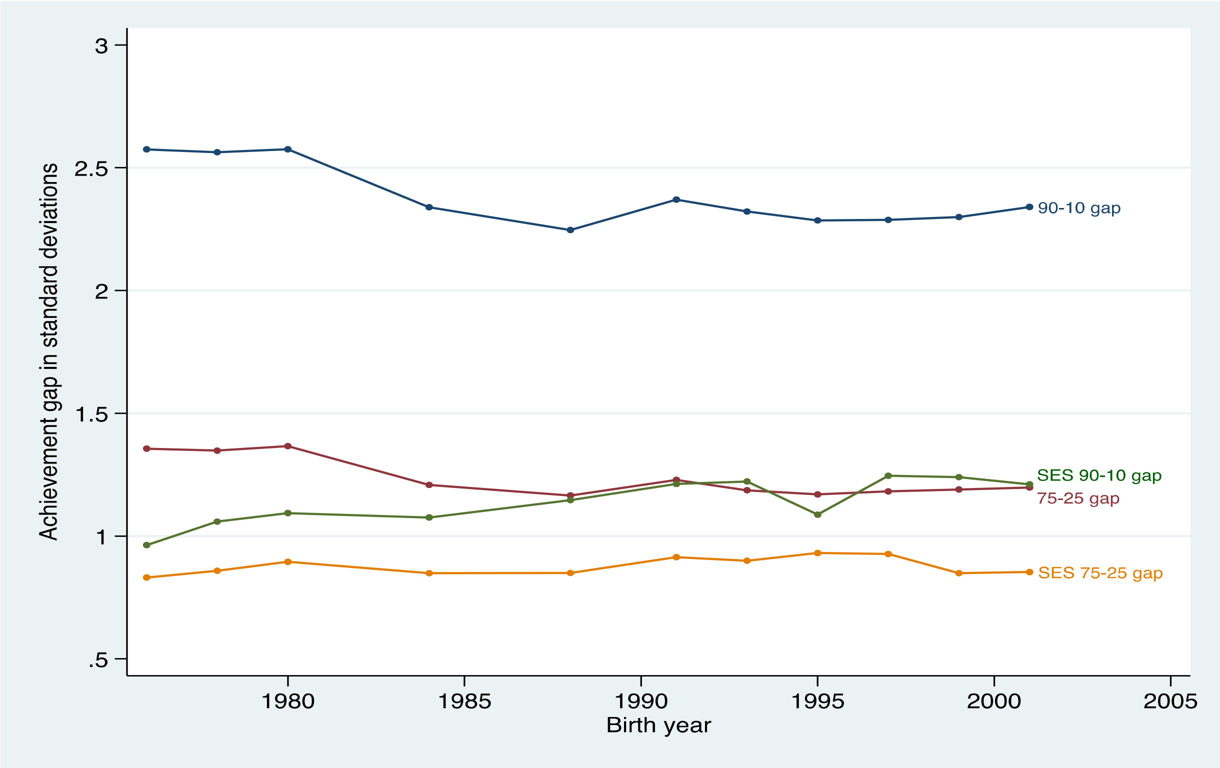
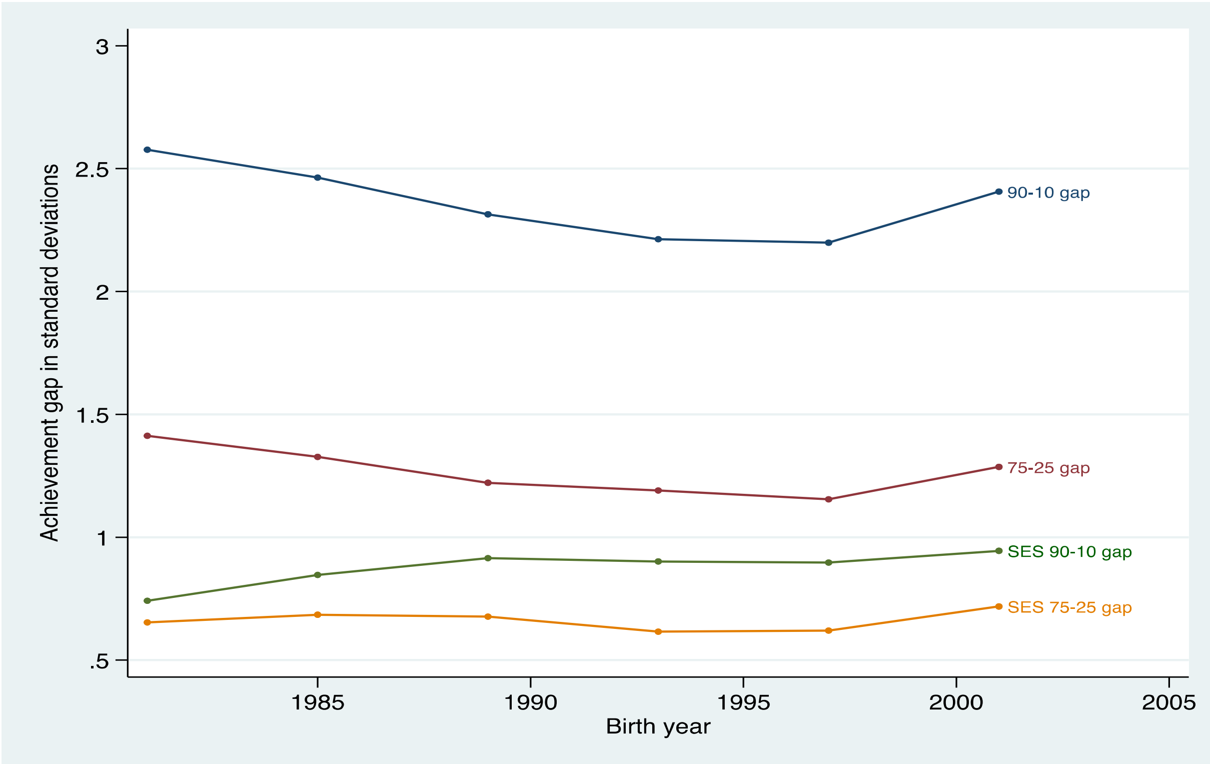


Figure C.3: Unconditional and SES Achievement Gaps in TIMSS, Birth Cohorts 1981-2001

Panel A: 8th Grade Math



Panel B: 8th Grade Science

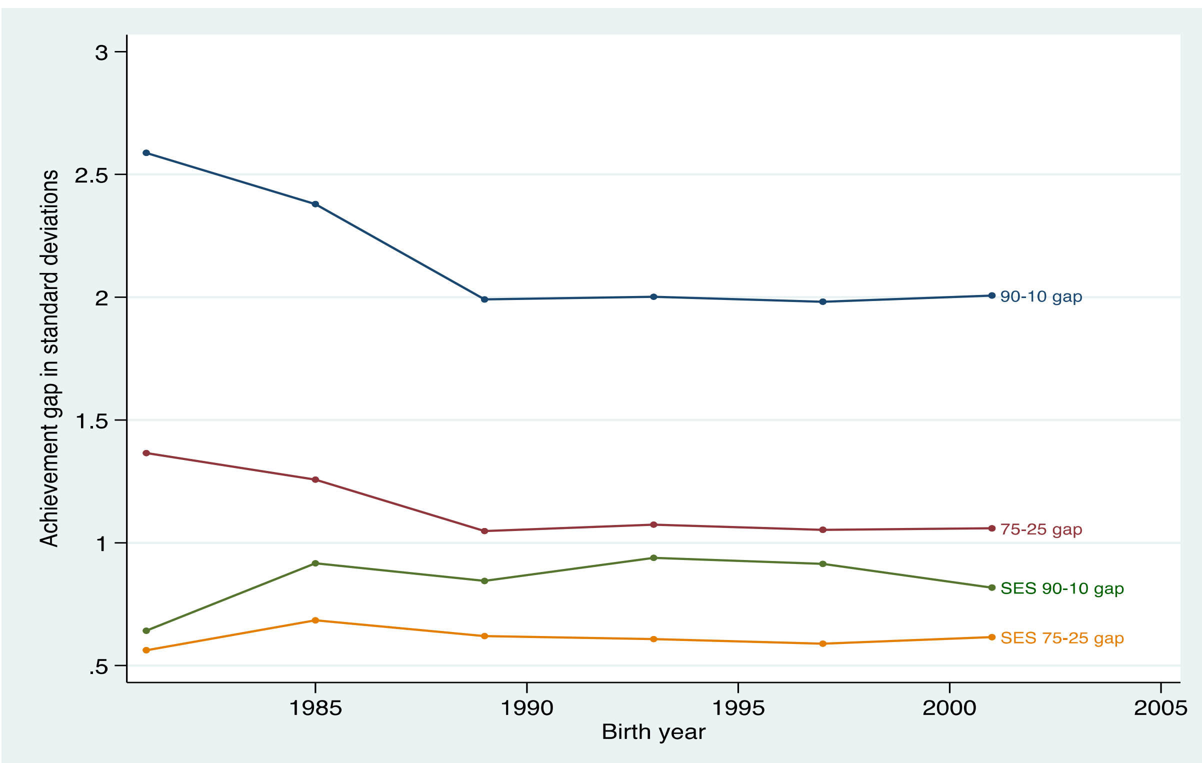
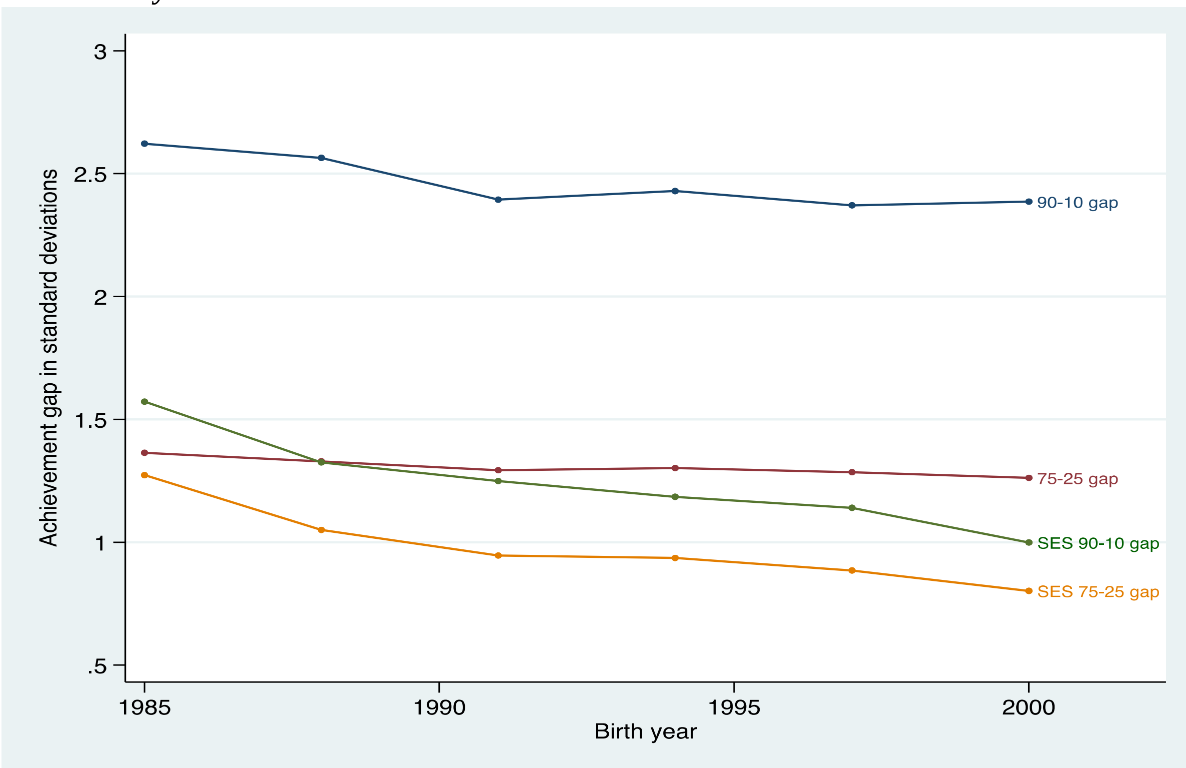
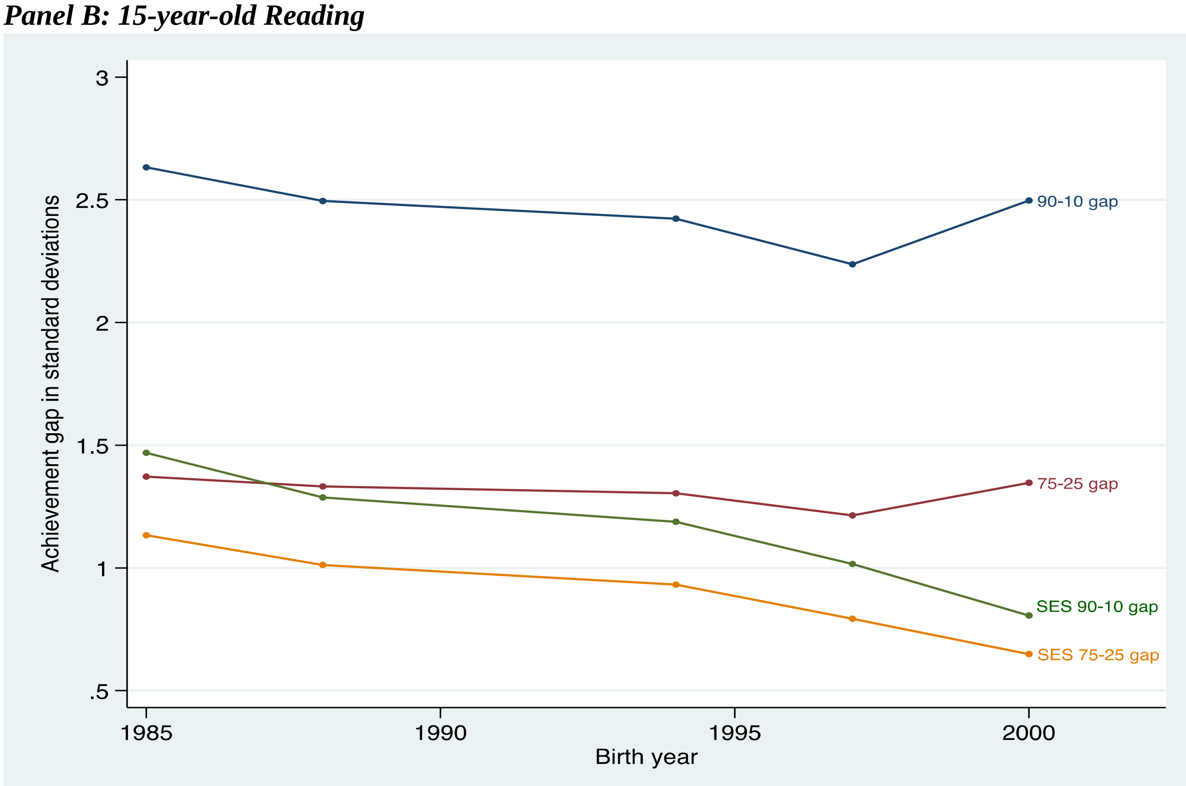


Figure C.4: Unconditional and SES Achievement Gaps in PISA, Birth Cohorts 1985-2000

Panel A: 15-year-old Math



Panel B: 15-year-old Reading



Panel C: 15-year-old Science

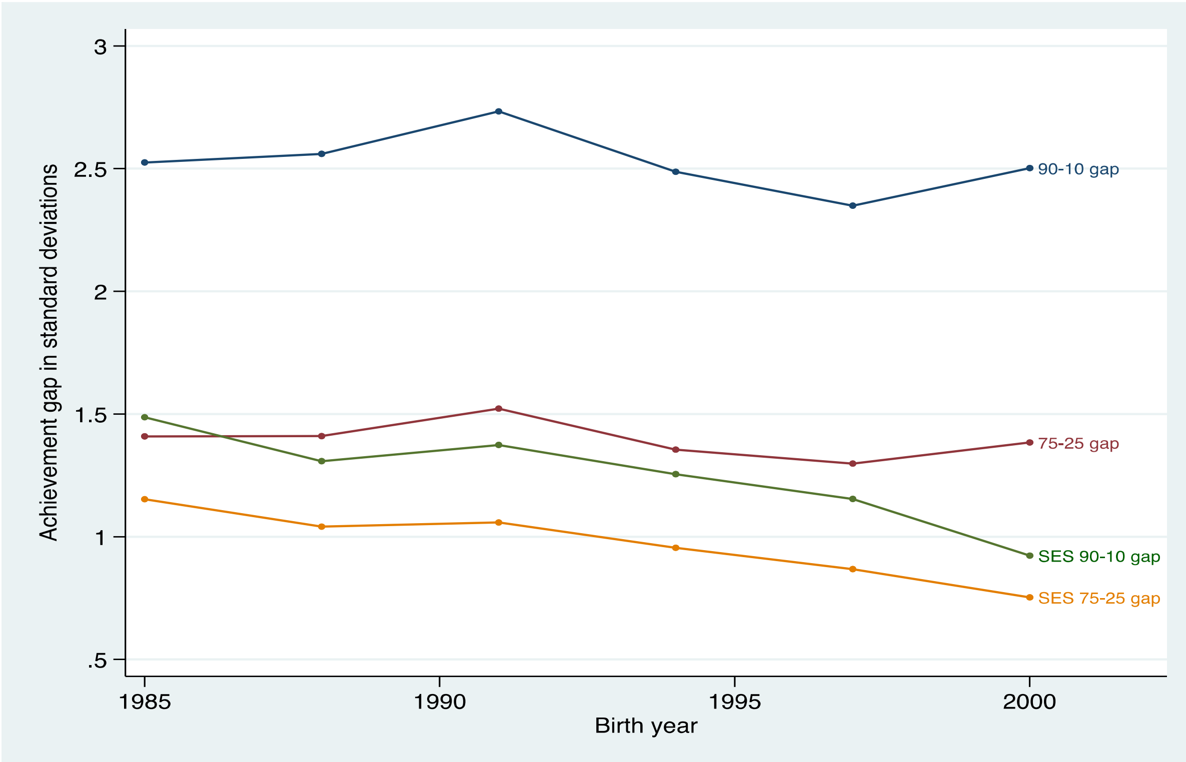
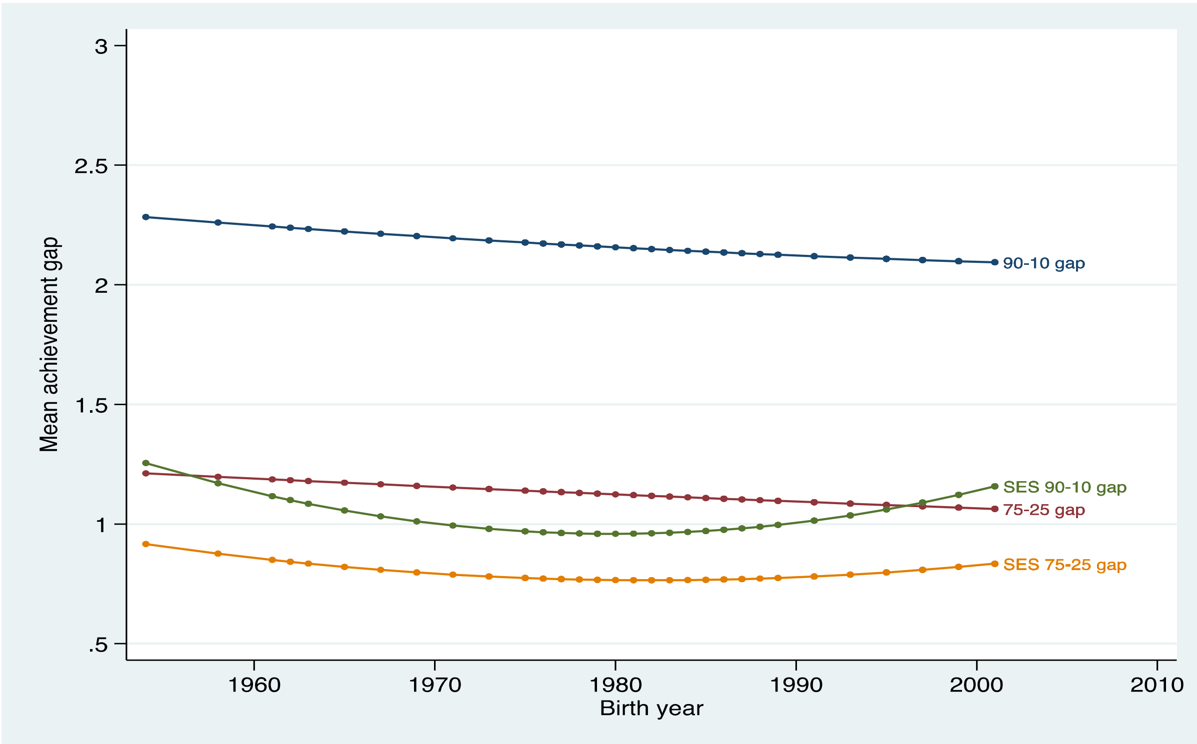


Figure D.1: Unconditional and SES Achievement Gaps Excluding PISA, Birth Cohorts 1954-2001



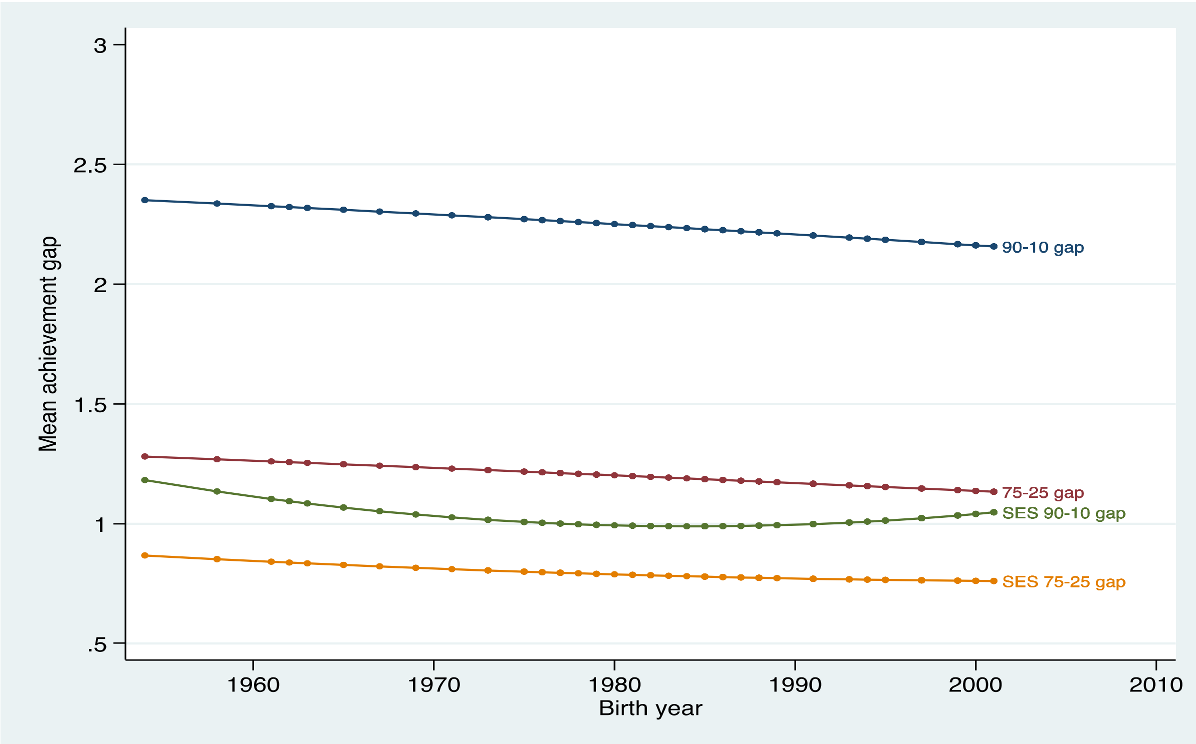
Note: See Figure 1 for data and methods. This figure is identical to Figure 1 except that it excludes PISA data.

Table 1: Description of Achievement Data

Test	Age/ grade	Birth cohorts		Observations by Test and Subject			
				Math	Reading	Science	Total
LTT-NAEP	age 13	1958-1999	Waves:	12	12	–	24
			Students:	99,450	115,780		215,230
LTT-NAEP	age 17	1954-1995	Waves:	12	12	–	24
			Students:	88,740	108,450		197,190
Main-NAEP	grade 8	1977-2001	Waves:	10	11	–	21
			Students:	1,004,650	1,122,980		2,127,630
TIMSS	grade 8	1982-2001	Waves:	6	–	6	12
			Students:	57,032		57,032	114,064
PISA	age 15	1985-2000	Waves:	6	5	6	17
			Students:	29,125	25,225	29,119	83,469
Total			Waves:	46	40	12	98
			Students:	1,278,997	1,372,435	86,151	2,737,583

Note: LTT-NAEP math is first tested in 1973, as opposed to reading which starts in 1971. For the 1973 math, data are only available for achievement levels and not for achievement gaps. Sample sizes for NAEP data are rounded to the nearest 10.

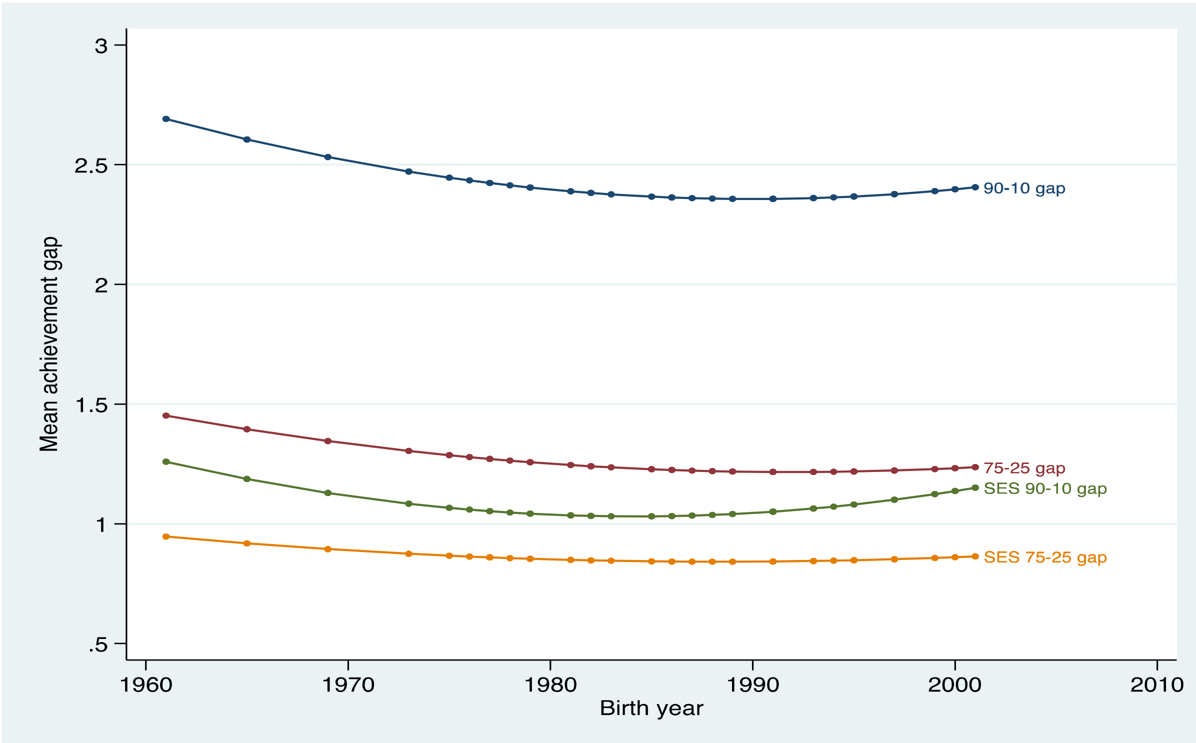
Figure 1: Unconditional and SES Achievement Gaps of U.S. Students, Birth Cohorts 1954-2001



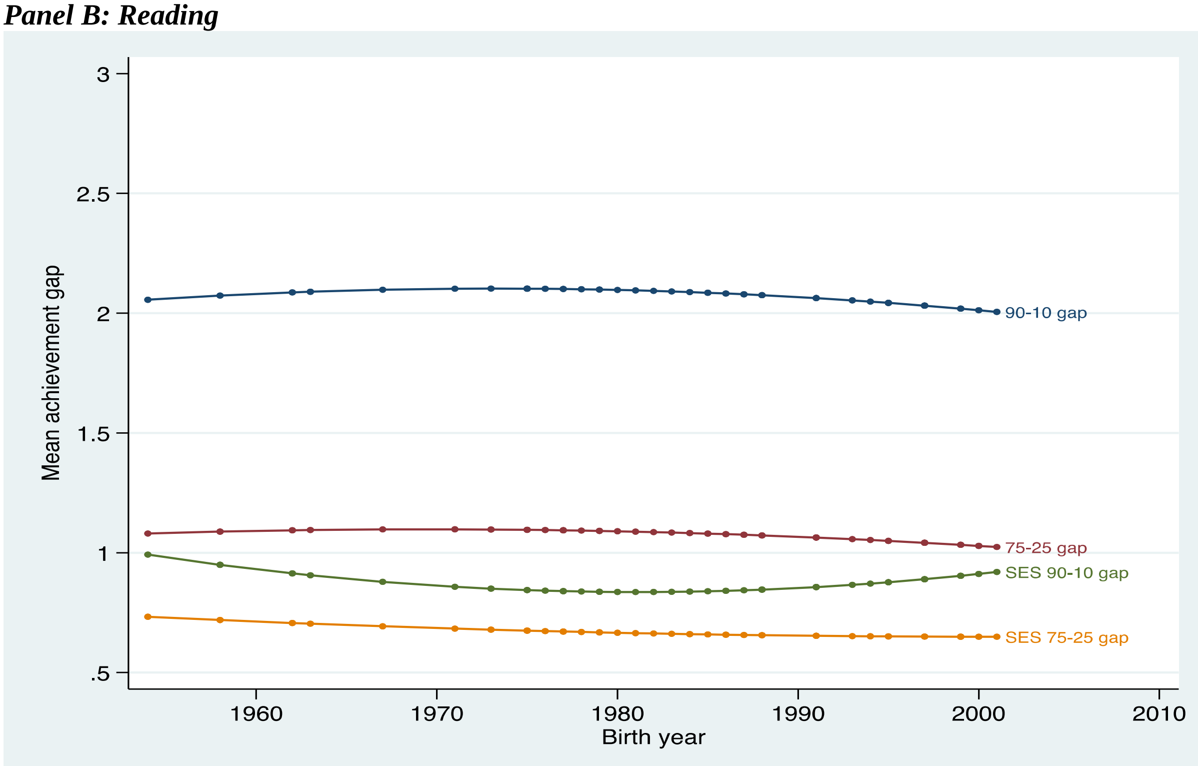
Note: All tests administered by LTT-NAEP, Main-NAEP, PISA, and TIMSS. 1954-2001 birth cohorts, all subjects, all students. 90-10 (75-25) gap: unconditional achievement difference between the students at the 90th and 10th percentiles (75th and 25th percentiles) of the achievement distribution. SES 90-10 (75-25) gap: achievement difference between the students in the top and bottom deciles (quartiles) of the SES distribution. Normalized achievement is measured in standard deviations. The s.d. presented is the difference between the year the test was administered and 2000 (or the closest year to that date available for a specific test series). Each marker indicates years where there are one or more underlying observations.

Figure 2: Unconditional and SES Achievement Gaps by Subject, Birth Cohorts 1954-2001

Panel A: Math

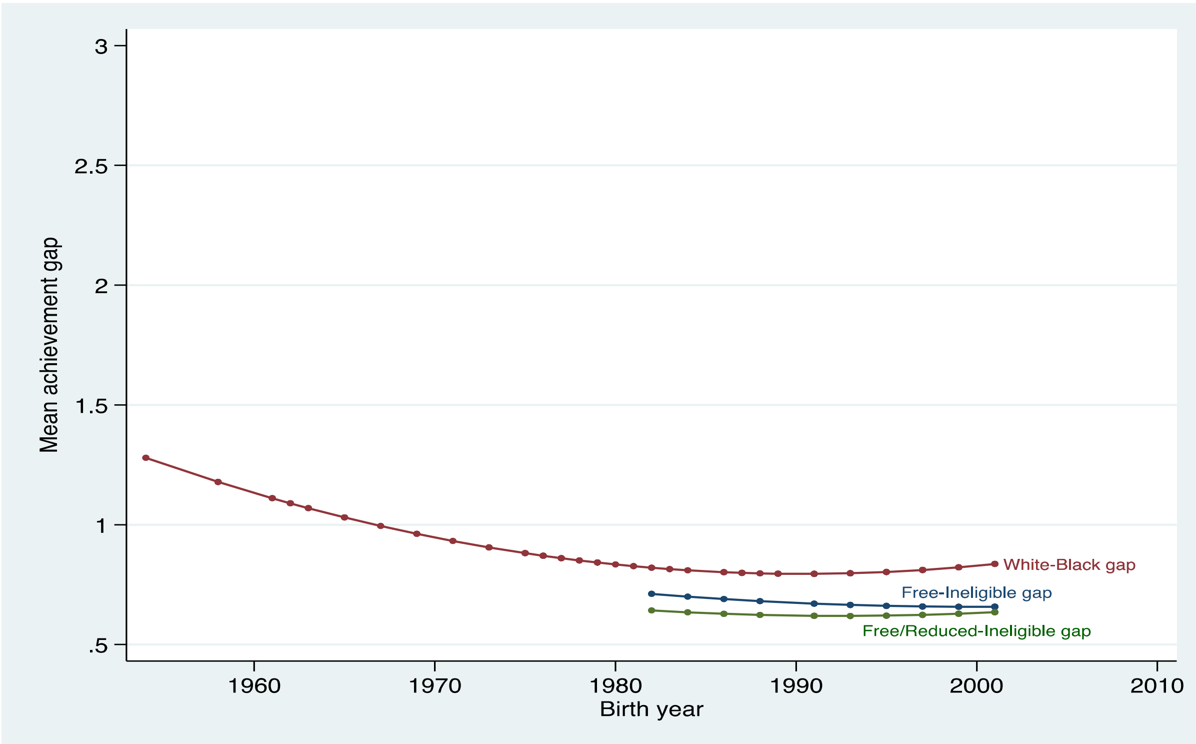


Panel B: Reading



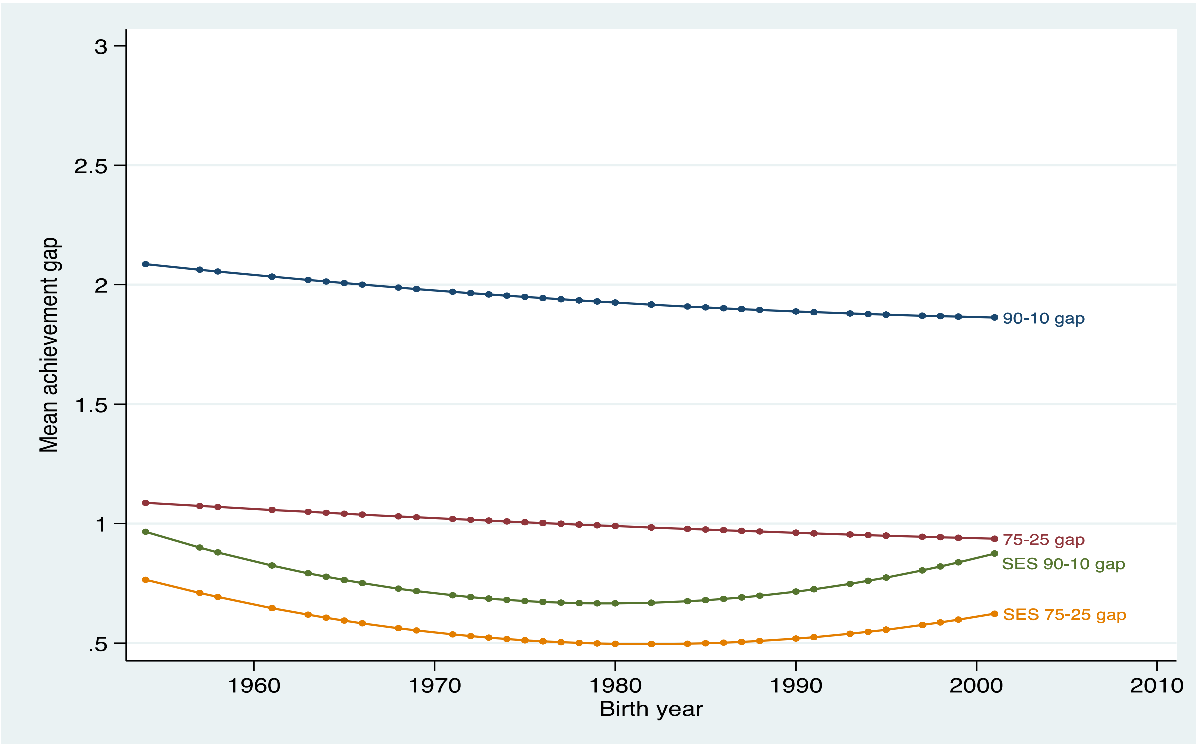
Note: See Figure 1 for data and methods.

Figure 3: Achievement Gaps for Eligibility for Free and Reduced-Price Lunch and for Race, Birth Cohorts 1954-2001



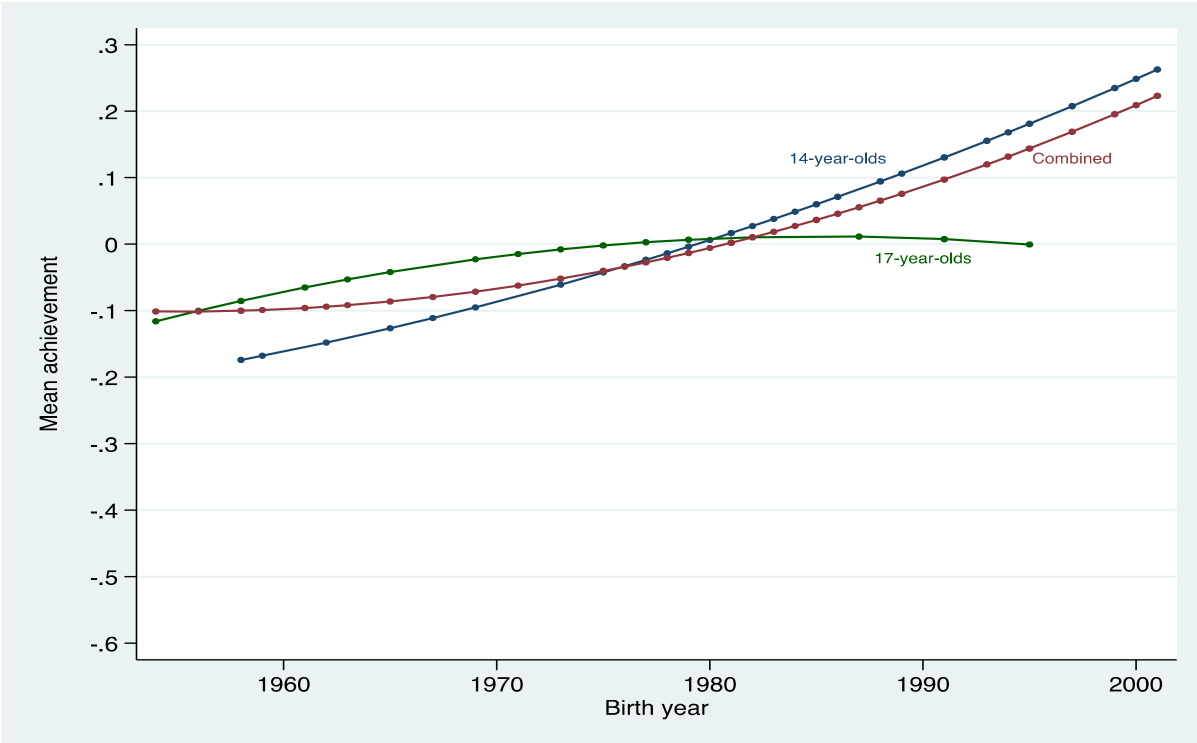
Note: Samples: For free and reduced price lunch, 1982-2001 birth cohorts, Main-NAEP surveys, math and reading, all students; for White-Black gap, 1954-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, black and white students. See Figure 1 for data and methods. Data on free and reduced-price lunch eligibility are only available for Main-NAEP tests, starting with the 1982 birth cohort.

Figure 4: Unconditional and SES Achievement Gaps among White Students, Birth Cohorts 1954-2001



Note: Sample: 1954-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, white students. See Figure 1 for data and methods.

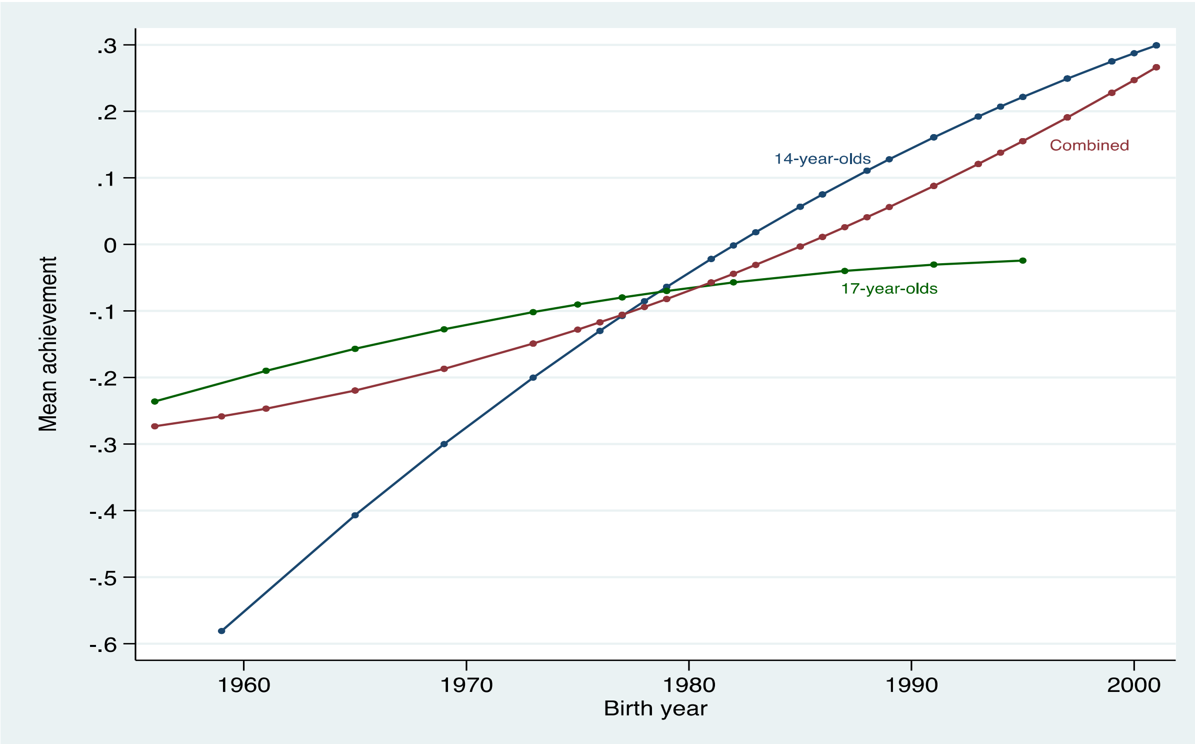
Figure 5: Achievement Levels of Younger and Older Students, Birth Cohorts 1954-2001



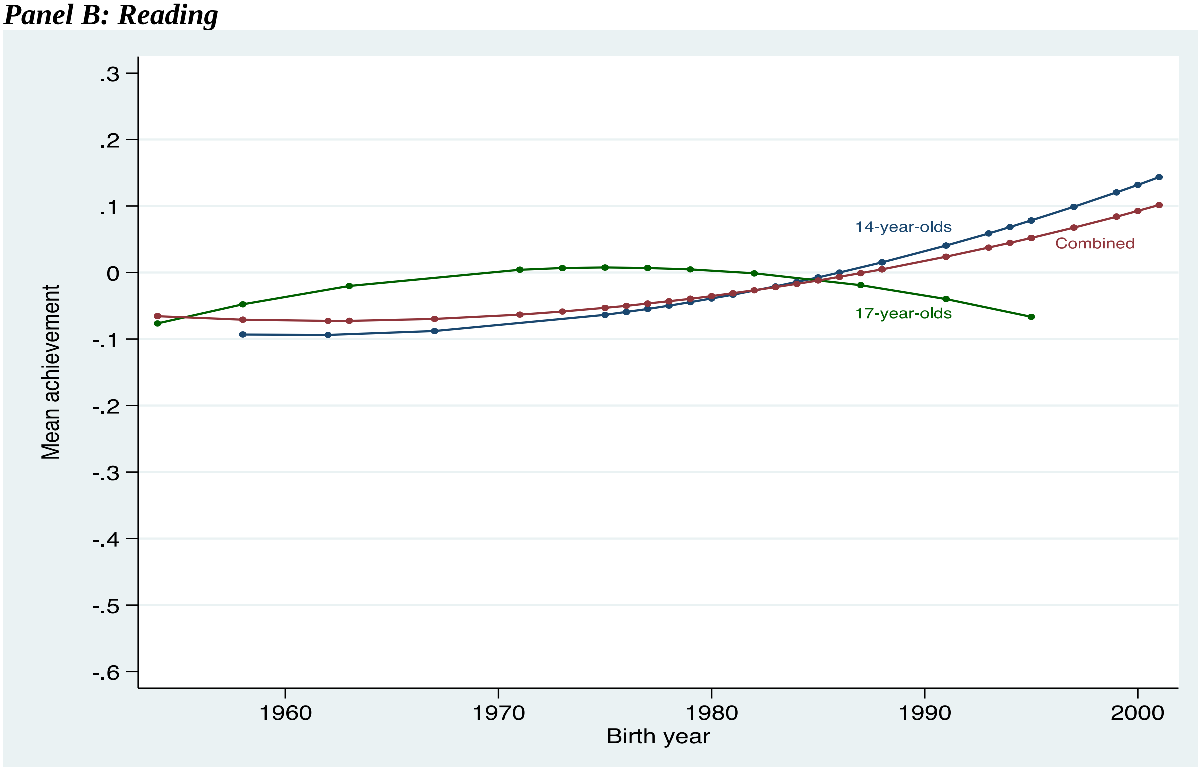
Note: Sample: 1954-2001 birth cohorts, all surveys, all subjects, all students. Younger students are those between ages 13 and 15 or in 8th grade, depending on the test. For expositional purposes, younger students are referred to as 14-year-olds. Older students are those aged 17 or in 12th grade, depending on the test. See Figure 1 for data and methods.

Figure 6: Achievement Levels of Younger and Older Students by Subject, Birth Cohorts 1954-2001

Panel A: Math



Panel B: Reading



Note: See Figures 1 and 5 for data and methods.