Incomplete Draft

Please Do Not Circulate

Closing the SES Achievement Gap: Trends in U.S. Student Performance

Eric A. Hanushek, Paul E. Peterson, M. Danish Shakeel, Laura M. Talpey, Ludger Woessmann[†]

September 11, 2019

Abstract

Concerns about 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 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 educational mitigating disparities.

[†] Hanushek: Hoover Institution/Stanford University, NBER, and CESifo; hanushek@stanford.edu. Peterson: Harvard University and Hoover Institution/Stanford University; ppeterso@gov.harvard.edu.; Shakee: Harvard University, danish_shakeel@hks.harvard.edu; Talpey: Hoover Institution/Stanford University; lmtalpey@stanford.edu. Woessmann: University of Munich, ifo Institute, and CESifo; woessmann@ifo.de. Helpful comments were received from Greg Duncan, Glenn Ellison, Magne Mogstad, Richard Murnane, Sean Reardon, Abhijeet Singh, Chris Taber, and participants at the IRP Summer Research Workshop and the CESifo Economic of Education conference.

In his first State of the Union Speech given in January 1964, President Lyndon Johnson declared a "war on poverty", saying "Our aim is not only to relieve the symptom of poverty, but to cure it and, above all, to prevent it ."[1] He then persuaded Congress to pass new educational programs designed to enhance the human capital of children born into poor and otherwise disadvantaged households. The rationale was simple and appealing: By providing the next generation with improved skills, the cycle of poverty could be broken in a structural way. Over the next half century, policies that focused on closing gaps in achievement between the rich and the poor were implemented and expanded by both the federal government and by state governments. This paper examines the overall outcomes of this set of policies by tracing the pattern of achievement gaps for disadvantaged and advantaged students since the inception of the War on Poverty. Our results show no relative improvements for the poor, indicating that future intergenerational mobility is unlikely to improve and that we need to reconsider policies and practices introduced to ameliorate the existing large achievement gaps.

It is well-documented that cognitive skills are directly related to economic outcomes in the labor market. Indeed, the U.S. rewards cognitive skills more than almost all other developed countries, which also implies that the U.S. more heavily punishes the lack of skills (Hanushek, Schwerdt, Wiederhold, and Woessmann (2015, (2017)). It is thus for very good economic reasons that Johnson and other policy makers have long searched for tools that will help schools break the linkage between a family's socio-economic status (SES) and students learning (Ladd (1996); Carneiro and Heckman (2003); Krueger (2003); Magnuson and Waldfogel (2008)). Without that, it is unlikely that the rate of intergenerational mobility will improve in the 21st Century.

Given the topic's importance, it is surprising that the pattern of SES-achievement gaps is so poorly documented. Popular commentary has linked widening income gaps in the United States to a perceived spread in the achievement gap between rich and poor. 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." 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." But the empirical basis for these conclusions is quite limited. The one systematic analysis supporting these conclusions is an influential study by Sean Reardon(2011), which finds widening achievement gaps by income after compiling data from multiple cross-sectional surveys.

We add to the relatively sparse literature by providing the first comprehensive analysis of recent trends in SES-achievement gaps, drawing upon data from four well-documented, intertemporally linked surveys of achievement in math, reading, and science. These assessments were administered to representative samples of cohorts of U.S. adolescent students who were born over the nearly five decades between 1954 and 2001. Each test was designed to be comparable across many years. This approach circumvents measurement problems encountered by non-linked cross-sectional data, permitting a

[1] https://en.wikipedia.org/wiki/War_on_Poverty [accessed August 31, 2019].

clearer picture of the differentials in performances of advantaged and disadvantaged students over the half century since the onset of the war on poverty.

Contrary to the popular commentary, we find little change in the SES-achievement relationship across the past fifty years. The SES-achievement gap between the top and bottom SES quartiles (75-25 SES gap) is roughly 0.9 standard deviations (s.d.), a gap implying a difference of three years of learning between the average student in the top and bottom quartiles of the distribution. And that gap has remained essentially unchanged for a half century, regardless of the attempts to ameliorate it with focused policies.

Moreover, that gap occurs within the context of stagnant levels of achievement overall among students nearing the end of their secondary education. While steady average gains in student performance have been registered over the past half-century among students at middle school, they do not translate into gains at age 17, the point students are expected to be ready for college and careers.

This paper has eight sections. The first reviews the literature on achievement gaps. The second and third describes our achievement data and SES measures. The fourth provides our methodological approach. The fifth displays trends in student achievement gaps and levels, and the sixth conducts a series of robustness checks. The seventh interprets these trends—or lack thereof. The final section concludes.

1. Relevant Research Literature

Definitions and measures of SES differ with context and data availability, but for the most part SES is "defined broadly as one's access to financial, social, cultural, and human capital resources," (National Center for Education Statistics (2012a)). Our focus is how the parents' SES relates to achievement of the child, which is taken as a proximate measure of future economic opportunities. Studies of intergenerational mobility compare parental SES and the child's SES as an adult, which is systematically related to the cognitive skills that the child takes to the labor market.

Here we first summarize the research that describes the relationship between SES of the family and children's outcomes. We then look in more detail at research that traces changes in that relationship over time.

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 linked to student achievement with school factors being much less significant. 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)). The literature is extensive enough that there have been a number of periodic reviews of the empirical relationship between SES and achievement (e.g., White (1982), Sirin (2005)).

One aspect of this extensive research is that limited attention has been given to the most appropriate measurement of SES. In empirical analyses data availability often dictates the choice of SES indicators. Nevertheless, for purposes of characterizing generational opportunities, this does not appear to be a significant problem because the commonly employed measures of adult education, household income, parental occupation, and the like tend to be highly correlated with each other and with general social and economic outcomes.

Reardon (2011) provides the most widely cited description of trends in SES-achievement gaps. The study has been frequently quoted in both academic literature and the general media,[2] and the conclusion that income-achievement gaps have dramatically increased over the past half century is arguably the contemporary conventional wisdom about SES patterns in achievement.

Reardon measures SES by survey data on household income. In analyses of data from 12 different cross-sectional surveys, he focuses on gaps in math and reading achievement of students at the 90th and the 10th percentile of the household income distribution.[3] 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). Interestingly, he finds that all of the change comes from changes in the 90-50 gap.[4] The 50-10 gap, arguably more relevant for assessing the war on poverty, remains unchanged over this period.

Three other trend studies, while focusing on international comparisons of SES-achievement gaps, include trends separately for the United States. Although each differs in its operational definition of SES, the data set it examines, and the specific methodology it employs, all three find a diminution on the size of the U.S. SES-achievement gap over the past two decades.

OECD (2018) estimates the change in the SES-achievement gap between 2000 and 2015, as traced through the PISA assessments, a consistent set of psychometrically-linked tests in math, science, and reading.[5] The OECD measure of SES is its index of "Economic, Social and Cultural Status (ESCS)," which aggregates data from students on their parents' education, their parents' occupation, and an inventory of items in their home. (This index is closely related to the measure used in our analysis and is described more thoroughly in section 3, below).

Instead of looking at achievement gaps for particular points on the SES distribution, it gauges changes in the SES-achievement pattern through changes in the "socio-economic gradient." Student performance on PISA is regressed on the ESCS index, and the amount of the variance explained (R^2) is interpreted as an indicator of the degree to which achievement is equitably distributed across the students in the

[2] For example, see Edsell (2012); Taverise (2012); Weissmann (2012); Maxie (2012); Duncan and Murnane (2014); Putnam (2015); and Jackson, Johnson, and Persico (2016).

[3] The comparisons of achievement gaps over time rely on first converting the gaps on the separate tests to standard deviations and assuming that these are comparable, but there is no discussion of this issue. Psychometricians have variously been concerned about how to interpret achievement that comes from tests with different scales (see, for example, Lord (1950), Dorans (2008), Ho (2009)). Hedges and Nowell (1998) also report SES-achievement correlations from a variety of cross-sectional studies. Using data from six surveys conducted between 1965 and 1992, test performance is regressed on parent education, income, and other variables. They find no clear trend in either the education or income coefficients. They conclude that changes in parental education and income do not explain patterns in the black-white achievement gap.

[4] As we discuss in Appendix D, this conclusion about the growth in the income-achievement gap is very fragile and highly dependent on the inclusion or exclusion of specific error prone data sets.

[5] PISA is the Programme for International Student Assessment; see Appendix A and http://www.oecd.org/pisa/.

survey. If more of the variance in performance is explained by the ESCS index over time, educational inequity is judged to increase. A falling R^2 is taken as a sign of improvement of the SES-achievement distribution. OECD (2018) reports a decline in R^2 over the fifteen-year period for the United States, indicating greater equity in the distribution of achievement.

Chmielewski (2019) combines data from one hundred countries on international tests conducted between 1964 and 2015 in order to estimate 90-10 SES-achievement gaps in all countries. She relies chiefly upon parental education as her primary SES indicator, although she also separately analyzes parental occupation and books in the home. She finds no significant trend for the United States on the eight test administrations of student performance given to cohorts born between 1950 and 2001. There are, however, concerns (discussed in the next section) about her treatment of tests that are not psychometrically-linked and about her reliance on just compressed education categories to measure SES.

In the third analysis, Broer, Bai, and Fonseca (2019) employ TIMSS assessments in math and science to estimate trends in SES-achievement gaps for eleven countries including the United States between 1995 and 2015.[6] They estimate 75-25 gaps on an SES index constructed from indicators of parent education, books in the home, and the presence of two education resources (computer and study desk).[7] They find that U.S. SES achievement gaps decline significantly for science performance but do not change significantly for math.

In addition to these direct investigations of SES-achievement gaps, a large number of studies look at the black-white test-score gap in the United States; see, for example, Grissmer, Kirby, Berends, and Williamson (1994), Grissmer, Flanagan, and Williamson (1998), Jencks and Phillips (1998), Magnuson and Waldfogel (2008), and Reardon (2011). These studies quite consistently identify a substantial closing of the racial test-score gap for cohorts born between 1954 and 1972, 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.

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 has been declining (Wilson (1987, (2011, (2012)) and black students constitute only around 16 percent of the school-age population (Rivkin (2016)). Thus, patterns of the black-white gap can only provide a limited picture of changes in the SES-achievement gap.

[6] TIMSS is the Trends in International Mathematics and Science Study and is conducted by the International Association for the Evaluation of Educational Achievement (IEA). See Appendix A and https://timssandpirls.bc.edu/.

[7] While more home resources are available, those employed were restricted to these two in order to maintain comparability over time and across countries. In the end, the distribution of their SES* index was computed separately for each country-year observation. Their index takes on a limited number of discrete values, and they fill in the achievement of the top and bottom quartiles by randomly sampling achievement values from the adjacent categories.

2. Longitudinal Achievement Data

Building on these studies, we estimate SES-achievement trends from four psychometrically-linked batteries of tests that span a 47-year time period. Each of these surveys uses a consistent data collection procedure to estimate the test performances of representative samples of U.S. adolescents over multiple years. Each test is designed to have a common scale and to be comparable over time by employing psychometric linkage based on using test items that are repeated across test waves.[8] All are 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 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.[9] 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 across 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.

Main National Assessment of Educational Progress (Main-NAEP)

Main-NAEP administers tests of math and reading aligned to the curriculum in grade 8.[10] Begun in 1990 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.[11] Main-NAEP maintains a reputation for reliability and validity similar to LTT-NAEP, and it was thought to

[8] As described below, however, the analysis employs fixed effects for tests and subjects and thus relies just on the patterns of achievement gaps within each test and does not use differences in scales across tests.

[9] 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 math 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.

[10] 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.

[11] 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.

track trends over time accurately enough that the LTT-NAEP no longer needed. For each administration of the test, the Main-NAEP sample is over 150,000 observations, the large sample being 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.[12] 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 and other sample details for each of the separate assessments are found in Appendix Table A.1.

3. Characterizing the SES Distribution

As noted, variation in data availability has generated a variety of alternative indicators of SES in prior investigations of the interplay of SES and achievement. The most commonly constructed indicators are drawn from the classic literature on social mobility – parental education, household income, parental occupations, and home possessions (National Center for Education Statistics (2012a), Sirin (2005)).

Our analysis is also dictated by the availability of information in the four assessment surveys included in this analysis. All four surveys provide student-generated information on parental education and a varying set of home possessions. The PISA survey also has student reports of parents' occupations. Following now-common practice, none of the surveys asked students to guess their parents' income.

We construct an SES index similar to the one used by OECD (2017a) that was mentioned above. Since the set of measured home possessions varies over time as does their individual utility for characterizing SES differences, we extract a principal component from a factor analysis separately for each test administration (for details, see Appendix B). These calculations assume, as does most equity and social

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

mobility research, that it is the relative position of the child's family in the SES distribution that is important to capture.

Because of data limitations, we depart from the OECD's measure of SES by not including occupational prestige, an item that is unavailable from TIMSS or either NAEP survey. 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 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).[13]

There are three underlying aspects of the survey data that lead to severe empirical complications in analyzing the pattern of SES-achievement gaps. First, for efficient responses and ease of coding, survey questionnaires rely heavily on categorical responses. Second, there are clear limits on the depth and breadth of any set of questions, leading typical survey designs to provide the most detailed information near the middle of the distribution of the population and not the very ends. This leads to potential categorical answers that do not provide great detail at the extremes, especially at the top end of the distribution. Third, the categorical background questions necessarily lead to categorical elements of the constructed SES distribution, and potential analytical problems induced by the limited categorization become severe at the top end of the SES distribution.

As a general rule, to be useful for estimating gaps between groups at the tails of the SES distributions, the SES construct must include items that can distinguish between those in each tail of the distribution and those who are not. Otherwise, serious measurement bias may be introduced by including observations that are outside the portion of the distribution under investigation.

In our data sets, the potential for this type of measurement error resulting from limited categorization of the SES distribution is particularly severe for some administrations of the Main NAEP survey, although it comes up also in the LTT NAEP. For example, in 1990, the first year for Main NAEP, students were asked about only four parental education levels and four items in their home. (See Table 1 for specific survey items). As a consequence, when we construct our SES index based on these two measures, 26.5 percent of the students that Main NAEP tested in 1990 were identified as in the top SES category. Similar problems emerge for other surveys in early administrations of both the Main NAEP and LTT NAEP.

As Table 1 also shows, there is a sharp contrast with the PISA survey in 2015, when the survey inquired about seven educational categories and twenty-two items in the home. As a result, just one percent of the sample falls into the top SES category in the PISA 2015 survey. Figure 1 shows the frequency distribution for distinct categories identified along the SES continuum for the two surveys; the tall spike at the top end of the NAEP 1990 SES distribution represents the 26.5 percent of the sample falling in the

[13] 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 proxy for educational inputs of the family. 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).

top SES category that can be observed. The more granular description of the SES distribution that follows the more detailed information on education and home possessions is obvious in the PISA 2015 data.

Our analysis takes two steps to avoid measurement error arising from crude characterization of the SES distribution. First, we do not attempt to estimate a 90-10 gap or comparable extremes of the distribution, as inspection of the data reveals that 45 of the potential 96 observations in our combined data sets have no way of distinguishing between students in the top 10 percent of the distribution and those further into the distribution. In other words, in close to half of the observations the top observed category of the SES distribution has greater than 10 percent of the students. Second, for our main analysis of the 75-25 SES gap (difference between students in the top and bottom quartiles of the SES distribution), we exclude all test administrations where we cannot isolate the SES group that falls in the relevant part of the distribution. In other words, if the highest SES category for a particular test administration includes more than 25 percent of the population, we exclude that test administration from our estimation of the trend in the 75-25 gap. This sample rule reduces our observations to 81 out of the potential 96 observations, but we are more confident about being able to identify achievement patterns in the tails of the distribution.[14]

The categorical nature of the SES-achievement distribution also means that we do not precisely observe the cut points that we use, such as the 75th percentile. For this, we interpolate achievement between the SES categories immediately above and immediately below the desired cut point.[15]

One approach that has been previously used is to describe achievement at different points of the SES distribution is to estimate regression functions for the SES-achievement relationship within the observed SES values and then to extrapolate the estimated relationship into the bottom and top portions of the distribution where there are no observed values (Reardon (2011), Chmielewski and Reardon (2016), Chmielewski (2019)).[16] This extrapolation actually involves two steps. In the first, under a linearity assumption, the data point for the extreme value of the SES-achievement distribution is filled in by assuming that the average achievement of the top group corresponds to the midpoint of the category. For example, in the 26.5 percent of the 1990 NAEP distribution in the top SES category, a data point would be created for the average achievement of the top group corresponding to the 13.25 percentile of the distribution. The second step estimates a cubic or linear regression function through the observed data points for a specific test-subject-year observation.

Figure 2 shows the relevant SES percentile-achievement data for Main NAEP in 1990 and the fitted cubic and linear regressions following this extrapolation procedure. These projections show considerable error where the data points diverge from the estimated line (particularly in the center of the

[14] In the empirical work, we investigate the sensitivity of the results to the sample reduction by also looking at the 70-30 gap that allows for 91 observations. This does not affect our overall results.

[15] Here we use local linearization. In dealing with the categorical SES distribution, Broer, Bai, and Fonseca (2019) used an alternate approach of randomly drawing the appropriate number of observations from the next lower SES category.

[16] In each of these cases, the objective is to estimate the difference in achievement between the highest and lowest deciles of the SES distribution (90-10 gap). Reardon (2011) and Chmielewski (2019) modify the projections if greater than 20 percent of observations were in the top or bottom quintile, relying on linear projections or adding dummy variables in the gap analysis.

distribution) and where the alternative projection lines diverge in the projection area of the plot. This latter issue with extrapolation is unfortunate because that is just the part of the plot that comes into play in the analysis of achievement gaps. In other words, the choice of extrapolation approach, about which there is no clear way to choose, directly interacts with the calculations of SES achievement gaps.

The measurement error problem is directly related to the range of possible categories of the SES distribution that are observed, and it will generally be more severe when the SES distribution is constructed from a single variable. For example, when Chmielewski (2019)) estimates the SES-achievement gap from parental information in the 2015 PISA data, she is estimating a 90-10 gap with a top category for parental education that includes 46 percent of the distribution.[17] In fact, the top education category averages over 40 percent of the test observations across our entire sample of different assessments at varying times – implying it would be necessary to extrapolate out of range *all* of the achievement points when looking at either 75-25 or 90-10 SES gaps based solely on surveyed parental education. Similarly within our data, the lowest education category includes 19 percent of the education distribution in the early years of the LTT NAEP, implying that it is sometimes also necessary to extrapolate at the bottom of the distribution. It is difficult to estimate precisely how large the extrapolation error might be, but it is clear that the measurement problem is most severe in the tails of the distribution – precisely the portion of the distribution that is the focus of any gap study.

An additional feature of our analysis is a focus on grouped SES-achievement comparisons. We consider the average performance of students whose families fall in the top quartile of the SES distribution compared to those in the bottom quartile. By focusing on this comparison, we do not have to characterize the precise pattern of achievement within the extremes of the SES distribution. The alternative approach (found in Reardon (2011), Chmielewski and Reardon (2016), Chmielewski (2019)) is to compare achievement at specific points in the SES distributions, namely the estimated achievement of somebody exactly at the 10th percentile or the 90th percentile. In order to make this point comparison, it is necessary to evaluate the expected achievement from the regression function, and this requires imposing a specific functional form on the tail distribution and relying heavily on the achievement patterns in the central part of the distribution to describe outcomes in the tails.

4. Methodological Approach

We compile an aggregate distribution of achievement from student-level microdata available for each subject, testing age, and birth cohort for close to 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 2 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.

[17] It also applies to the use of parental income, because surveys that include income measures invariably ask about income in fixed categories. Again, the survey categories rarely provide detail on the upper portions of the income distribution. See Appendix D.

The Main-NAEP and TIMSS tests are grade based, while the LTT-NAEP and PISA tests are administered to students at specific 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 standard deviations (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.[18]

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 tested 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 includes indicators for subject, age group, and administrative entity.

For the trends in performance levels, we calculate the mean performance, $\bar{O}_{isa}^{t}$, 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:

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

where δ_i, γ_s, and λ_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 average student performance for two groups, *j* and *k*, where the gap at any time *t* is Δ_{jk}^{t}. Specifically, we estimate:

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

In our main analysis, we begin with 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 quartiles of the SES index distribution.[19] For expositional purposes, we refer to this as the 75-25 SES gap.

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

[19] As noted, focusing on the 90th and 10th percentiles has been popular in past work, but both LTT NAEP and Main NAEP lack observations at the 90th percentile and above. Using our criteria of dropping observations if no data are found in the relevant category, 12 out of 13 years of Main NAEP would be dropped, as would 11 of 30 years for LTT NAEP.

As noted, the gaps for the SES distributions are calculated as the average score for those above the 75th SES percentile and the average score for those below the 25th percentile, allowing us to compare achievement across disparate groups without imposing any functional relationship on the SES-achievement distribution.[20] By doing this, we can make useful comparisons of achievement for different broad segments of the population, but we do not need to impose any functional relationship on the SES-achievement distribution.

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.

Analyzing 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 both black and white students.

[20] The categorical distribution of SES scores leads us to interpolate the achievement between the SES points just above and just below the 25th and 75th percentiles to obtain the relevant values (see Appendix B). An alternative is to estimate the SES-achievement relationship (such as shown in Figure 2 and then to calculate the achievement at the 25th and 75th SES percentiles predicted by the estimated relationship. Doing this, however, has minimal results on the patterns of SES gaps, as shown below – as long as the calculations are confined to the SES percentiles actually observed in the data.

5. 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 alternative portions of the SES distribution.

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

5.1 Unconditional Achievement Disparities

We begin by considering changes in the overall distribution of achievement. In Figure 3, we plot the unconditional gaps measured in initial standard deviations for the 90-10 and the 75-25 gaps over the past half century.[21] 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.[22] 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, almost 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).

5.2 Achievement Disparities by SES

The underlying SES-achievement gap trend for in the 75-25 SES-achievement gap in Figure 4 presents a startling picture: The connection between SES and achievement hardly wavers over the entire period. In the 1961 birth cohort (he first where the 75-25 gap is reliably observed), the achievement gap between the average of those in the top and bottom quartiles of the SES distribution was 0.84 s.d. As the trend line in Figure 4 indicates, this gap increases to 0.91 s.d. for the cohort born in 2001.

This trend line is based on the within-test data and does not use the between-test data. Each of the within-test trends, as seen from the scatter of data points, parallels the overall trend with the exception

---

[21] Note that it is possible to get good estimates of the bottom and top of the achievement distribution, because the unconditional performance distribution does not rely on characterizing the SES of parents.

[22] 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.

of the PISA trend. The PISA gaps, as previously indicated in the OECD (2018) analysis, shows a closing of SES gaps over the observed birth cohorts since 1985. If estimated from the 64 observations excluding the PISA data, the overall trend shows a statistically significant bow with the predicted 1961 birth cohort gap of 0.86 rising to 0.99 for the 2001 birth cohort. However, we know of no reason why the PISA gaps are less informative or less reliable than those for the other three tests, and thus we know of no reason to exclude the PISA data. If we exclude each of the other tests one at a time and re-estimate the trend, we see flat, straight lines with the joint quadratic parameters being insignificantly different from zero for each subsample.

The estimation of the 75-25 gaps is based on the 81 observations that provide reliable information on achievement for the top quartile of achievement. If we expand the tails slightly to look at the 70-30 distribution, we have 91 (of the potential of 96) observations reliably observed. But Figure 5 makes it clear that, while the gaps are necessarily slightly less than for the interquartile range, the pattern of no change holds for this expanded sample. Indeed both individually and jointly it is not possible to reject the null that the coefficients for the linear and quadratic terms in the trend equation are zero for either the 75-25 SES gaps or the 70-30 SES gaps[23]. Figure 5 also indicates that we fill in more years of data for the 70-30 SES trend, but the earliest years still do not have reliable information for the top end of the SES distribution and the gaps begin with the 1961 birth cohort. (Markers on the trend lines indicate years of observations).

Trends are quite similar for math and reading separately (Figure 6). The 75-25 gap for math narrows slightly over time. In reading, the pattern increases slightly over the entire period. In other words, the aggregate trend of Figure 4 is not masking different trends across subjects.

Figures in Appendix C display plots of unconditional gaps and 75-25 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. There is some instability in the SES gap estimates for the individual assessments by subject. The individual figures simply connect available data points even if various intermediate points are missing because of inability to observe data for the 75th percentile and above. Overall, however, the four underlying assessments produce patterns of the SES gaps that are all similar with the exception of the general downward trend of the PISA gaps that was noted above.

There is the possibility, however, that changes occur in various parts of the distribution, perhaps offsetting each other. Figure 6 compares the top quartile with the bottom half of the distribution (75-50 gap) and the top half of the distribution with the bottom quartile (50-25 gaps). While there appears to be a slight downward curvature to the 50-25 gaps, the estimated quadratic coefficients for both gaps are again insignificantly different from zero.

Overall we find flat trends in SES-achievement gaps. This finding can be viewed from two different vantage points. We do not find a narrowing in gaps despite still being in the midst of a half-century "war on poverty." Nor do we find any confirmation of the finding of Reardon (2011) that there have been large increases in gaps at the top end of the distribution. As described in Appendix D, his results appear to reflect a reliance upon error-prone cross-sectional data and the inclusion of surveys that were particularly unsuited for analyzing trends in gaps.

[23] For the 75-25 gaps, $F(2, 71) = 0.53$; for the 70-30 gaps, $F(2, 82)=1.05$.

5.3 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.[24] As can be seen in Figure 8, 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 (those eligible for reduced price lunch), 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 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 this income-achievement gap reveals only miniscule change over the course of two decades.

We do not find this binary measure of family income to be the most appropriate way to look at trends, particularly given the changing definition of eligibility discussed previously. Nonetheless, these results are entirely consistent with the trends for the 75-25 SES-achievement gaps.

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 8. 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.[25]

---

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

[25] 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).

To see whether trends in achievement gaps are driven by shifts in ethnic composition, we estimate the SES-achievement gap for both white and black students separately (Figure 9). These calculations use the overall national SES distributions for both (as used in the aggregate 75-25 SES analysis above). Interestingly, the pattern of SES gaps for both whites and blacks is virtually identical. The overall black-white gaps are of course influenced by very different positioning of the races in the SES distribution. As Table 3 shows, both whites and blacks have moved up in the SES distribution between 1978 and 2012. At the same time, blacks have gained in relative terms, and this shows up particularly in the representation in the bottom 25 percent of the distribution.

In general, however, it is difficult to argue that changes in the ethnic composition of student cohorts account for the unwavering SES-achievement gap.

5.4 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 10 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 (with appropriate fixed effects). 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.[26] 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. In other words, 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).[27] 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.[28]

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.

---

[26] As noted, students labelled as age 14 actually span 13-15 years old.

[27] 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.

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

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 11a). 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 11b).[29] 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)).

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.

5.5 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 above the 75th percentile of the SES distribution are around three years ahead of those below the 25th percentile by 8th grade.

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.

6. Further Robustness Checks

Our approach differs from other related work in a variety of ways, so it is useful to investigate whether some of the analytical choices lead to the differing conclusions.

[29] 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.

6.1 Point Estimation v. Group Calculations

Because of the limited information about achievement in the upper tail of the SES distribution, we focused on the 75-25 SES gap, and we compared performance in ranges at the tails of SES levels. The alternative is to focus on gaps at points in the distribution. We do this in two ways. First, keeping our samples that are defined by having reliable information about performance in the tails, we can still take achievement in a narrower band around the median. Specifically, instead of taking the top half of the distribution or the bottom half of the distribution, we can calculate achievement for students in the 45-55 percentile band and then consider the 75-50 gap or the 50-25 gap using this revised median achievement. In these alternatives, there is a bend upward in the 75-50 gap – increasing by about 0.15 s.d. between the 1961 birth cohort and the 2001 birth cohort with most of the gain coming after the 1970 birth cohort. There is a similar rise for the 50-25 gap (with the redefined median measure). Neither materially changes the conclusions above.

The second approach estimates a cubic function (like that in Figure 2) for each test-subject-year group and uses that to define the relevant achievement-SES gaps. This approach, following the methodology of Reardon (2011) and Chmielewski (2019), is used to expand to all of the available data and to create other gaps – particularly a 90-10 gap. While this approach has little effect on the estimated 75-25 gaps, it has a more noticeable impact on the estimated 90-10. In the latter case, there is approximately a 0.20 s.d. increase in the gap over the past 25-30 years. While not as large as found in Reardon (2011), the curvature of the gaps with a fairly steep recent period is striking. The issue of gap extrapolation appears to contribute to the different conclusions between Reardon (2011) and us.

6.2 Ordinal Achievement Data

There does exist a very different perspective on how to analyze and interpret achievement data of the kind used here. Various researchers have focused on the possibility of over-interpreting differences in achievement from standardized tests. Ho (2009), for example, evaluates potential shortcomings of comparing performance on the basis of standard deviations, as is done here, and how that relates to test scales and to different tests. Nielsen (2015) focuses on the fact that common achievement tests provide only ordinal, or rank, data. Thus, calculating the average achievement levels and comparing standard deviations of differences is suspect. This latter argument is also pursued by Bond and Lang (2013) who provide an analysis of the radically different inferences that can be drawn with the same (rank) information for blacks and whites where different order-preserving scale transformations are used.

One direct way to consider how the distribution has changed in an ordinal way is to construct "Lorenz-like curves" that trace out the inequality between achievement for the students in the 75th percentile of the achievement distribution as compared to those in the 25th percentile. Students within the bottom and top quartiles of the SES distribution obviously have varying performance on the tests. To compare the different performance of the SES groups, we calculate the score achieved at each percentile of students in the bottom quartile of the SES distribution, and then see the percentile of students in the top quartile that score at or below that score. As with a Lorenz curve, equal achievement distributions (i.e., no aggregate achievement gaps) would appear as a 45-degree line when we plot the percentage distribution of the low SES group against the corresponding percentages of the high SES group. This percentile-percentile static curve for any test-subject can then be observed over time. To do this effectively, we need the dense SES distributions that come from the PISA and TIMSS tests. It is then possible to compare the most recent test distribution with prior distributions using just the rank information in the tests.

Figure 12 compares the PISA math distributions for the 75-25 SES percentiles and for the 50-25 SES percentiles in 2015 and in 2000. In the top panel it can be seen that the 2000 distribution moved to being more equitable (closer to the 45-degree line). In the bottom panel, we also see that the 50-25 has moved very similarly. In sum, the distribution of performance has become more equitable, and this is seen across the distribution. This pattern, it turns out, is precisely what we saw for the individual PISA results and what is easy to see in Appendix Figure c9 for the size of the PISA math gaps in standard deviations across test years.

Figure 13 does the symmetric comparison for the TIMSS math distributions in 1995 and 2015 (which relate to the comparisons in s.d. in Appendix Figure c7). Interestingly, the TIMSS distribution for the 75-25 comparisons has become less equitable over time although the changes are less pronounced than the prior PISA changes. This is precisely what we found in the aggregate comparisons of the individual tests – PISA pointed to reductions in gaps while TIMSS did not.[30]

The ordinal comparisons are not easily compared across multiple time periods and across different tests. We also do not have readily available statistical tests when making these combined comparisons. In short, interpreting the composite trends across multiple assessments as we do here is not practical. But, at least from the straightforward bivariate tests with specific assessments, we conclude that treating the assessment data as ordinal rankings does not materially affect the conclusions here.

7. Discussion of Trends

The previous discussion has linked achievement to family socio-economic status, largely as motivated by ideas of intergenerational mobility. This perspective takes student achievement as being largely predictive of future economics outcomes and thus of how future economic wellbeing relates to parents' SES.

There is a different, more common perspective that comes from the extensive investigation of educational production functions. A simple educational production function model underlies much of the public and academic discourse, namely:

$$achievement = \text{ } family\ inputs + schools + other \tag{3}$$

In this, parents are a prime input into the education of children, and in empirical analysis SES of the family is taken a reliable proxy for the education in the home.

Our aggregate trend data cannot of course identify the causal effect of each component of this relationship. While previous studies generally involving micro-level studies of student achievement

[30] The TIMSS conclusions differ somewhat from those of Broer, Bai, and Fonseca (2019), which found math scores becoming more equal by SES group although they were not statistically different over time.

have provided credibly estimated relationships for individual factors under specific circumstances, there are few if any attempts to consider the aggregate impacts of the demographic shifts and policy thrusts.

The underlying motivation for this analysis was derived from a set of major policies designed to close achievement gaps at all levels of government. 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:

- 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 that began with the war on poverty and expanded state programming over time 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)).

The previous evidence, however, shows that these policies have not dented the serious gaps in achievement that follow the SES of parents.

One common response to this is simply "but the gaps would have grown were it not for these programs." In other words, if the family side of the equation has gotten worse, the programs might have been exactly balancing to these worsening trends.

It is difficult to cite definitive evidence on this. At the aggregate level, it is clear that there are more single parent families now than in the past. At the same time, the difference in age of first child by SES has grown. It is also true that the overall distribution of income has widened. Each of these might be expected to lower the educational inputs of families and to expand the class differences. But parents have also become more educated over time, a factor generally presumed to have a powerful positive influence on family education. Family sizes have gotten smaller, and the difference by SES has shrunk. And child poverty rates have remained quite constant.

Fundamentally, however, the measurement of SES in empirical studies reflects the specific data available in a particular analysis and not any judgment about the underlying structure of education in the home. The common interpretation of the various family measures employed and discussed above is that they are proxies for more structural features of family education that connect SES and achievement. But as such this does not lend itself to understanding trends in family inputs.[31] Nor do they lend themselves to policy interventions, because it is unclear whether changing the measured characteristics of families will also change the education in the home.

Even without making any strong statements about the balance of forces, however, it seems clear that the existing policy programs are not doing what we hoped they would. Thus, re-evaluating their design and focus seems appropriate.

The pattern of achievement levels is also an unresolved puzzle. While the different achievement trends for younger and older students has been noted previously, no satisfactory explanations have been found (Krueger (1998); Hanushek (1998)). 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. Blagg and Chingos (2016) consider four other 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 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.[32] 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. This age pattern remains unexplained.

[31] In their review, Cheng and Peterson (2018) provide illustrative possible mechanisms at work. 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)). 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)). These and other childhood or adolescent experiences undoubtedly contribute to SES disparities in academic achievement (Kao and Tienda (1998); Perna (2006); Goyette (2008); Jacob and Linkow (2011)).

[32] 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.

8. Conclusions

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.

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, science, and reading that has accumulated.

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 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 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 SES gaps. These unwavering gaps suggest reconsidering existing policy thrusts.

References

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

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.

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

Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results." *The Review of Economics and Statistics* 95, no. 5 (Decenber): 1468-1479.

Broer, Markus, Yifan Bai, and Frank Fonseca. 2019. *Socioeconomic Inequality and Educational Outcomes: Evidence from Twenty Years of TIMSS*: IEA Research for Education and Springer Open.

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.

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

Chmielewski, Anna K. 2019. "The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015." *American Sociological Review* 84, no. 3: 517-544.

Chmielewski, Anna K., and Sean F. Reardon. 2016. "Patterns of Cross-National Variation in the Association Between Income and Academic Achievement." *AERA Open* 2, no. 3 (July-September): 1-27.

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.

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.

Dorans, Neil J. 2008. *The Practice of Comparing Scores on Different Tests*. R&D Connections. Princeton, NJ: Educational Testing Service (August).

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.

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.

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.

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.

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

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., 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. 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.

Ho, Andrew Dean. 2009. "A Nonparametric Framework for Comparing Trends and Gaps Across Tests." *Journal of Educational and Behavioral Statistics* 34, no. 2 (June): 201-228.

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.

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

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.

Lord, Frederic M. 1950. *Notes on comparable scales for test scores*. ETS REesearch Report Series. Princeton, NJ: Educational Testing Service (July 13).

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. 2012a. *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation*. Washington, DC: National Center for Education Statistics (November).

National Center for Education Statistics. 2012b. *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.

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

OECD. 2018. *Equity in Education: Breaking Down Barriers to Social Mobility*. Paris: O.

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.

Sirin, Selcuk R. 2005. "Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research." *Review of Educational Research* 75, no. 3 (September): 417-453.

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.

White, Karl R. 1982. "The relation between socioeconomic status and academic achievement." *Psychological Bulletin* 91, no. 3: 461-481.

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 (2017)).

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 (2012b)). 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 (2017)). NAEP and TIMSS include similar questions about students' home possessions, but they do not provide a summary index.[33] 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.[34]

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.

[33] 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?"

[34] 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 (2017)). 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.[35]

[35] 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	
pisa	2000	15	math	493	2.622	1.364	
pisa	2003	15	math	482	2.564	1.329	
pisa	2006	15	math	474	2.394	1.293	
pisa	2009	15	math	488	2.429	1.302	
pisa	2012	15	math	481	2.371	1.285	
pisa	2015	15	math	470	2.386	1.262	
pisa	2000	15	reading	504	2.632	1.372	
pisa	2003	15	reading	494	2.495	1.332	
pisa	2009	15	reading	500	2.423	1.304	
pisa	2012	15	reading	497	2.237	1.214	
pisa	2015	15	reading	498	2.497	1.347	
pisa	2000	15	science	499	2.525	1.409	
pisa	2003	15	science	490	2.560	1.410	
pisa	2006	15	science	489	2.733	1.522	
pisa	2009	15	science	502	2.487	1.355	
pisa	2012	15	science	497	2.349	1.298	
pisa	2015	15	science	496	2.502	1.384	
timss	1995	14	math	487	2.577	1.413	
timss	1999	14	math	502	2.463	1.327	
timss	2003	14	math	504	2.314	1.222	
timss	2007	14	math	509	2.213	1.190	
timss	2011	14	math	509	2.199	1.154	
timss	2015	14	math	518	2.406	1.286	
timss	1995	14	science	521	2.588	1.365	
timss	1999	14	science	515	2.379	1.257	
timss	2003	14	science	527	1.991	1.048	
timss	2007	14	science	520	2.002	1.074	
timss	2011	14	science	524	1.982	1.053	
timss	2015	14	science	529	2.007	1.059	
naep	1990	14	math	259	2.589	1.427	
naep	1992	14	math	262	2.940	1.590	
naep	1996	14	math	271	2.785	1.487	
naep	2000	14	math	276	2.817	1.475	
naep	2005	14	math	279	2.774	1.454	
naep	2007	14	math	281	2.742	1.443	
naep	2009	14	math	283	2.776	1.456	
naep	2011	14	math	284	2.767	1.458	
naep	2013	14	math	285	2.771	1.460	
naep	2015	14	math	282	2.832	1.496	
naep	1990	14	reading	255	2.554	1.345	
naep	1992	14	reading	254	2.542	1.337	
naep	1994	14	reading	254	2.554	1.355	
naep	1998	14	reading	263	2.320	1.199	
naep	2002	14	reading	264	2.228	1.156	
naep	2005	14	reading	262	2.351	1.219	
naep	2007	14	reading	263	2.303	1.177	
naep	2009	14	reading	264	2.267	1.160	
naep	2011	14	reading	265	2.269	1.172	
naep	2013	14	reading	268	2.280	1.180	

Table A.2 (continued)

Test	Test year	Age	Subject	Mean	Unconditional gaps 90-10	Unconditional gaps 75-25	SES gaps
naep	2015	14	reading	265	2.321	1.188	
naepltt	1978	13	math	264	2.846	1.532	
naepltt	1982	13	math	268	2.442	1.298	
naepltt	1986	13	math	269	2.276	1.183	
naepltt	1990	13	math	270	2.314	1.208	
naepltt	1992	13	math	273	2.273	1.202	
naepltt	1994	13	math	275	2.372	1.236	
naepltt	1996	13	math	275	2.334	1.209	
naepltt	1999	13	math	275	2.415	1.256	
naepltt	2004	13	math	281	2.399	1.247	
naepltt	2008	13	math	281	2.470	1.253	
naepltt	2012	13	math	285	2.535	1.318	
naepltt	1971	13	reading	255	2.025	1.053	
naepltt	1975	13	reading	256	2.026	1.027	
naepltt	1980	13	reading	258	1.944	1.041	
naepltt	1988	13	reading	258	1.907	1.025	
naepltt	1990	13	reading	257	2.029	1.050	
naepltt	1992	13	reading	260	2.218	1.162	
naepltt	1994	13	reading	258	2.224	1.141	
naepltt	1996	13	reading	258	2.228	1.135	
naepltt	1999	13	reading	259	2.156	1.146	
naepltt	2004	13	reading	259	2.054	1.070	
naepltt	2008	13	reading	260	2.072	1.045	
naepltt	2012	13	reading	263	2.067	1.062	
naepltt	1973	14	math	266			
naepltt	1973	17	math	304			
naepltt	1978	17	math	300	2.580	1.408	
naepltt	1982	17	math	298	2.425	1.305	
naepltt	1986	17	math	302	2.305	1.251	
naepltt	1990	17	math	305	2.314	1.287	
naepltt	1992	17	math	307	2.215	1.196	
naepltt	1994	17	math	306	2.278	1.178	
naepltt	1996	17	math	307	2.264	1.203	
naepltt	1999	17	math	308	2.328	1.246	
naepltt	2004	17	math	307	2.172	1.152	
naepltt	2008	17	math	306	2.174	1.164	
naepltt	2012	17	math	306	2.246	1.169	
naepltt	1971	17	reading	286	2.536	1.327	
naepltt	1975	17	reading	285	2.447	1.278	
naepltt	1980	17	reading	285	2.335	1.232	
naepltt	1988	17	reading	290	2.110	1.115	
naepltt	1990	17	reading	290	2.298	1.202	
naepltt	1992	17	reading	290	2.429	1.232	
naepltt	1994	17	reading	288	2.466	1.290	
naepltt	1996	17	reading	287	2.339	1.220	
naepltt	1999	17	reading	288	2.349	1.215	
naepltt	2004	17	reading	285	2.445	1.234	
naepltt	2008	17	reading	286	2.484	1.268	
naepltt	2012	17	reading	287	2.372	1.229	

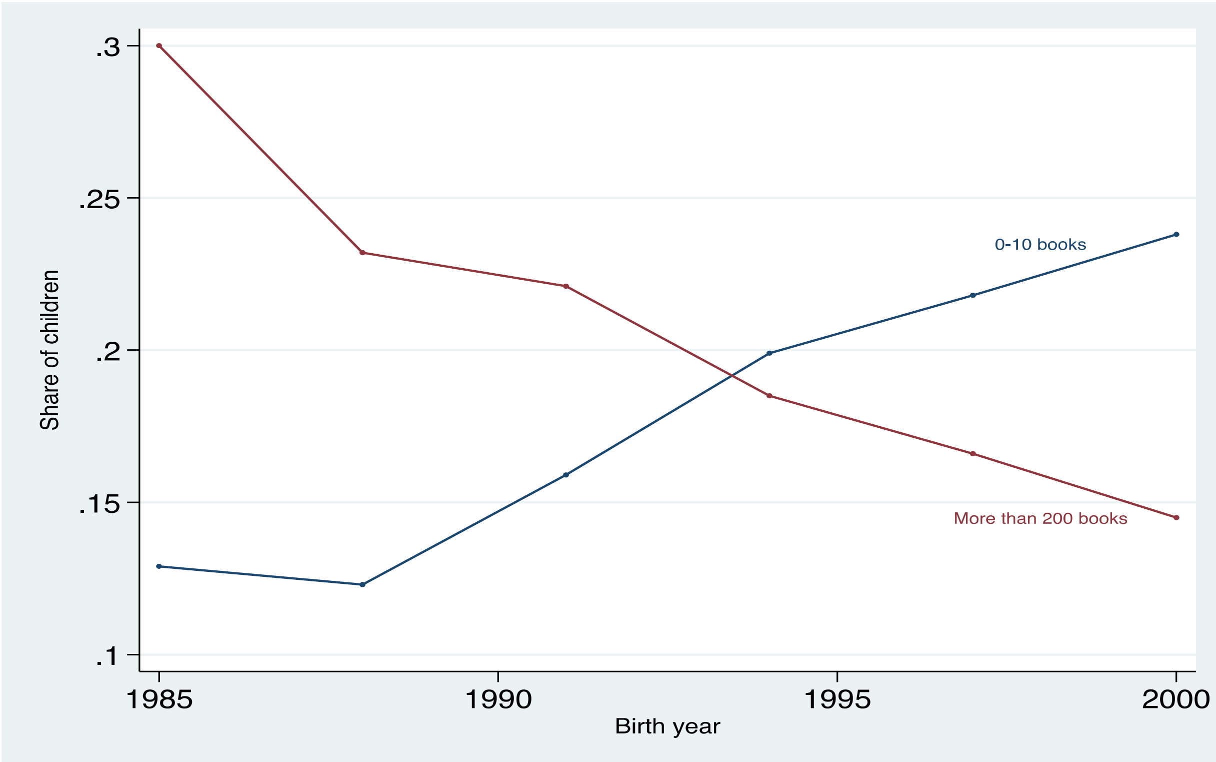
Table B.1: Correlation between SES Index and Family Income in the ELS

	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

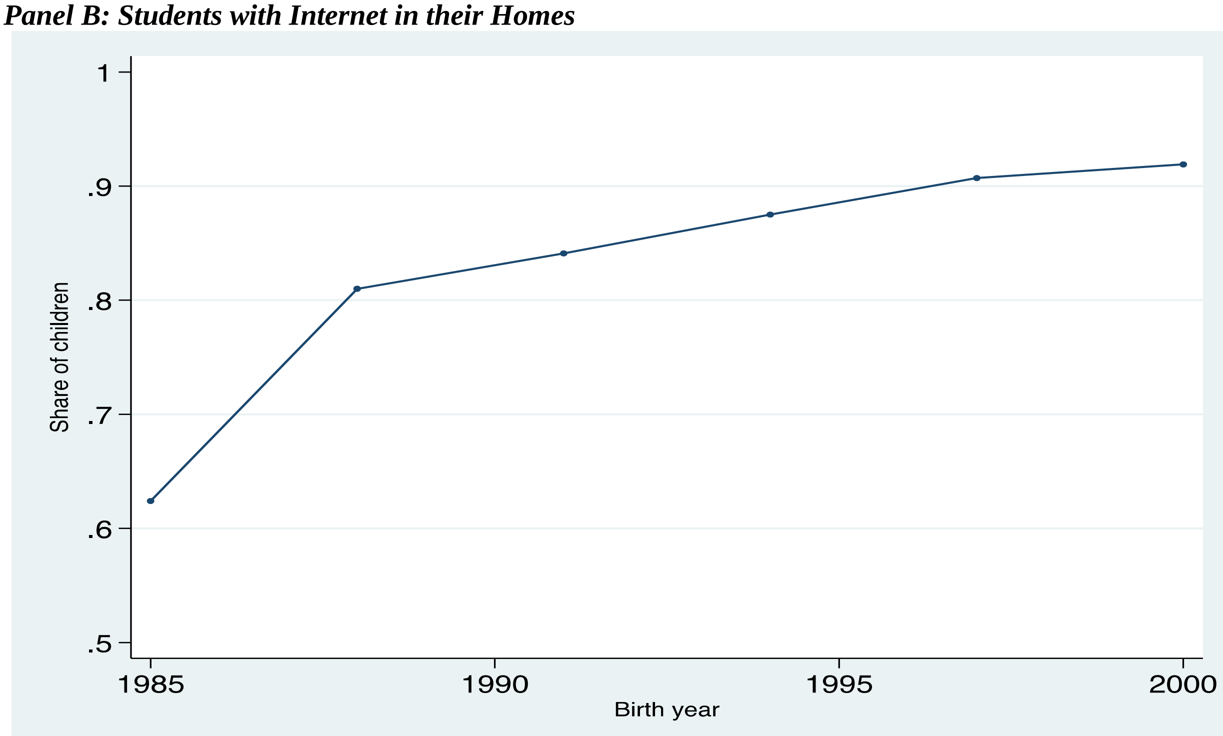
Note: Data source: 1988 Education Longitudinal Study (ELS).

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

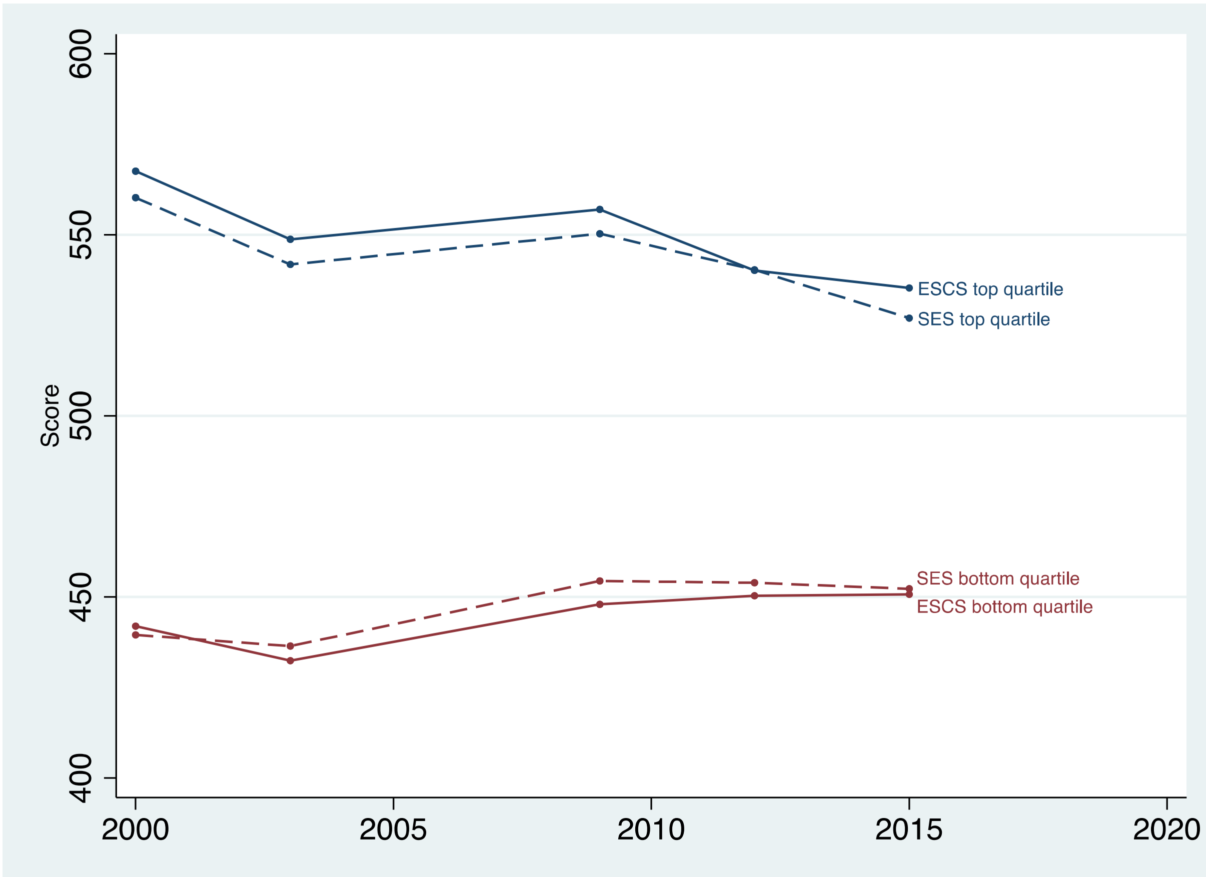


Panel B: Students with Internet in their Homes



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. Unconditional and 75-25 SES Gaps for Individual Assessments

Figure c1. NAEP Unconditional and SES 75-25 Gaps, Grade 8 Math

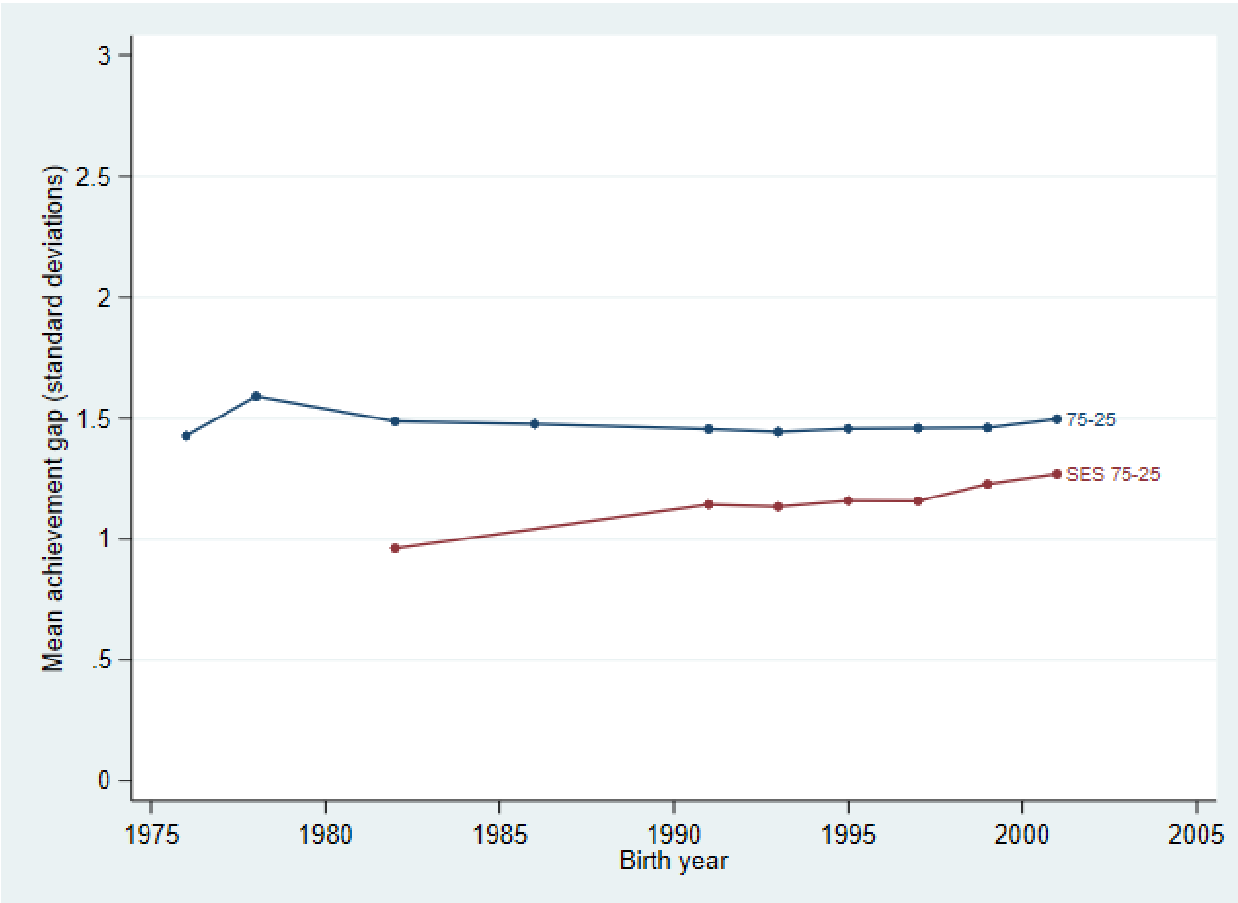


Figure c2. NAEP Unconditional and SES 75-25 Gaps, Grade 8 Reading

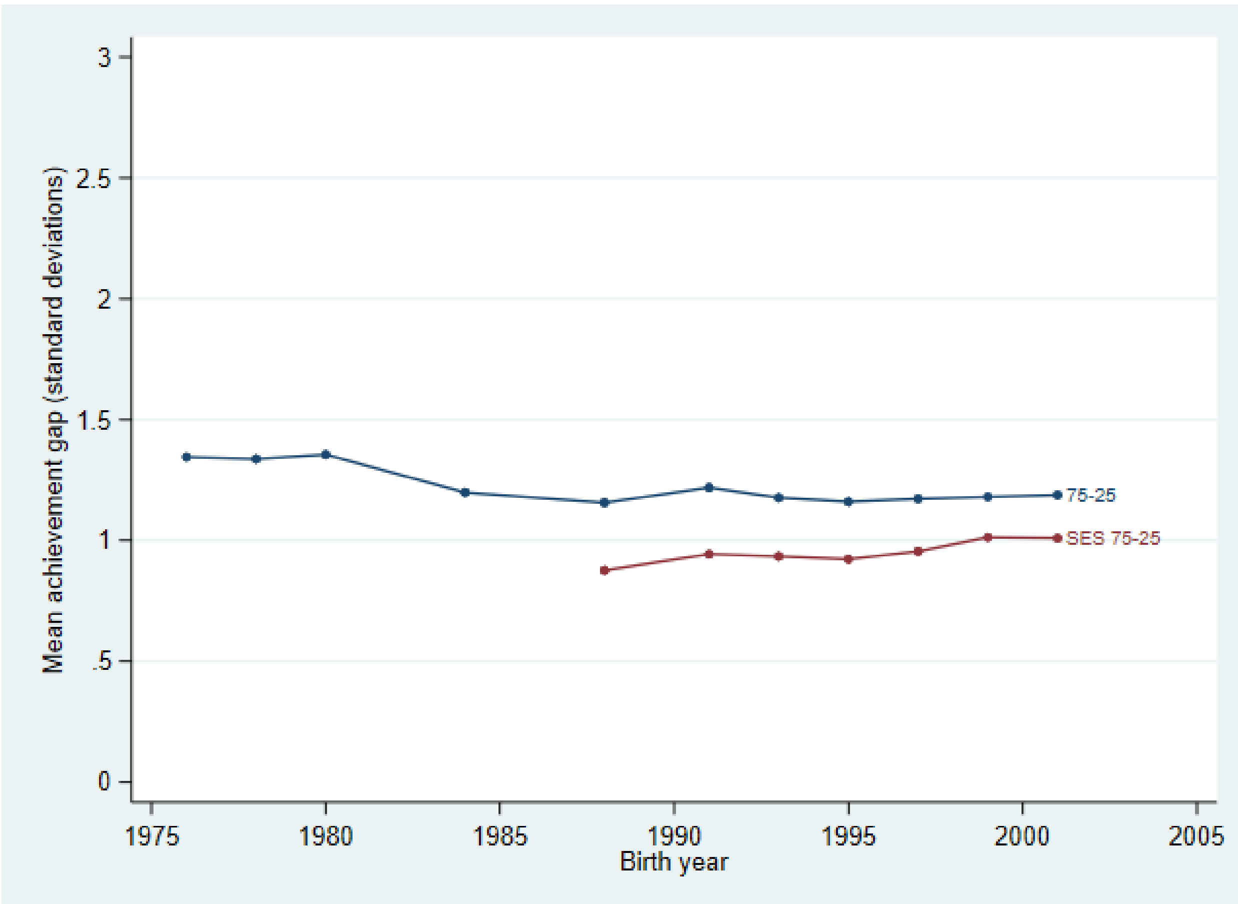


Figure c3. LTT-NAEP Unconditional and SES 75-25 Gaps, Age 13 Math

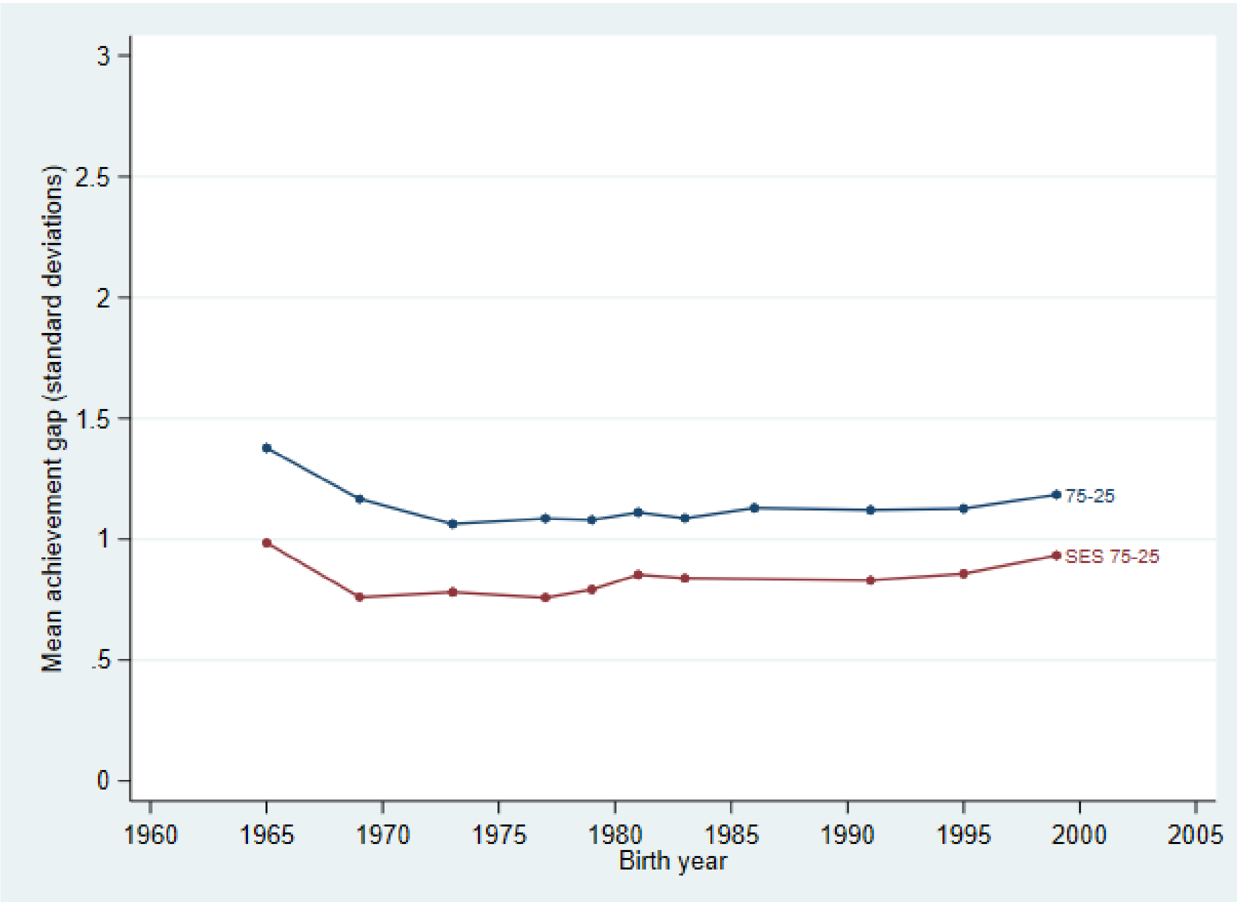


Figure c4. LTT-NAEP Unconditional and SES 75-25 Gaps, Age 13 Reading

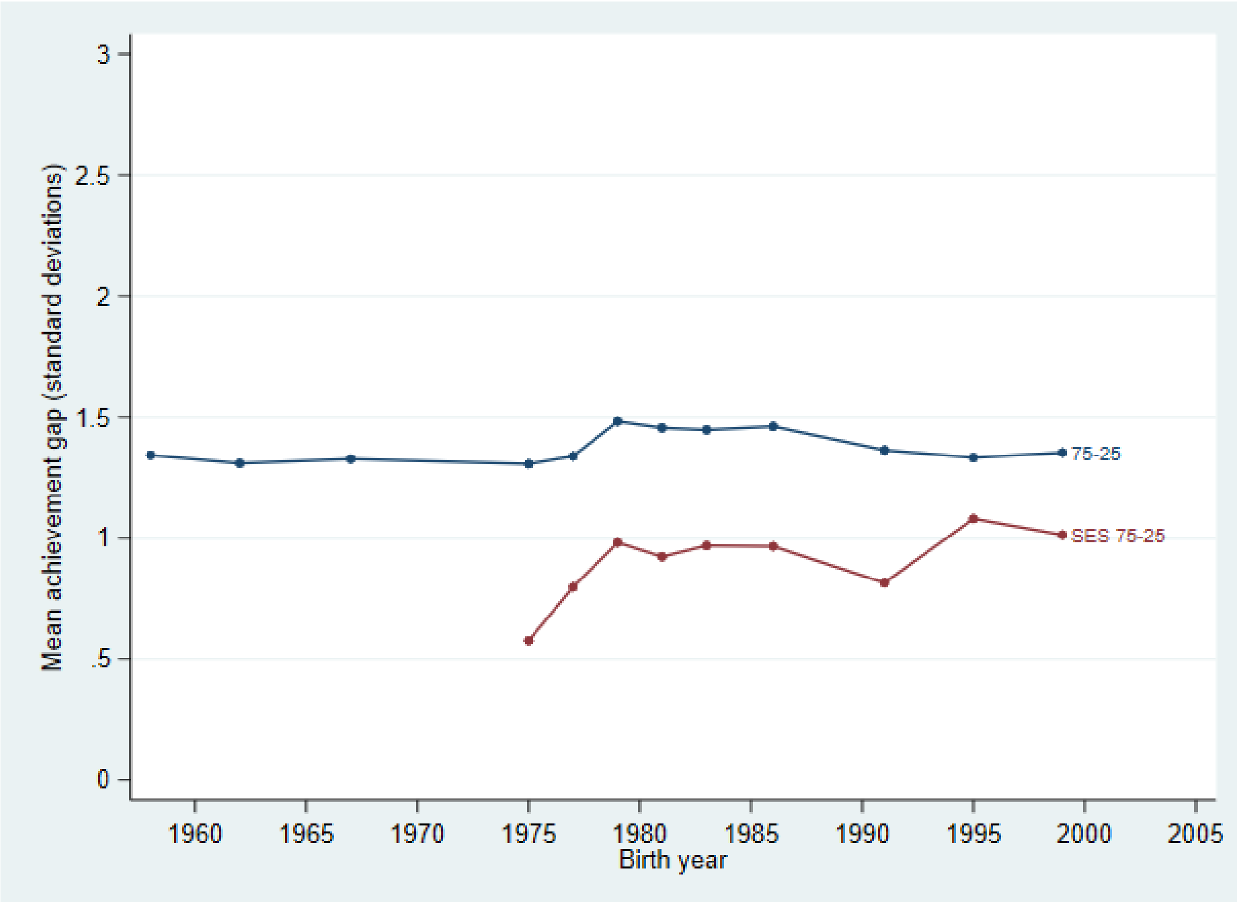


Figure c5. LTT-NAEP Unconditional and SES 75-25 Gaps, Age 17 Math

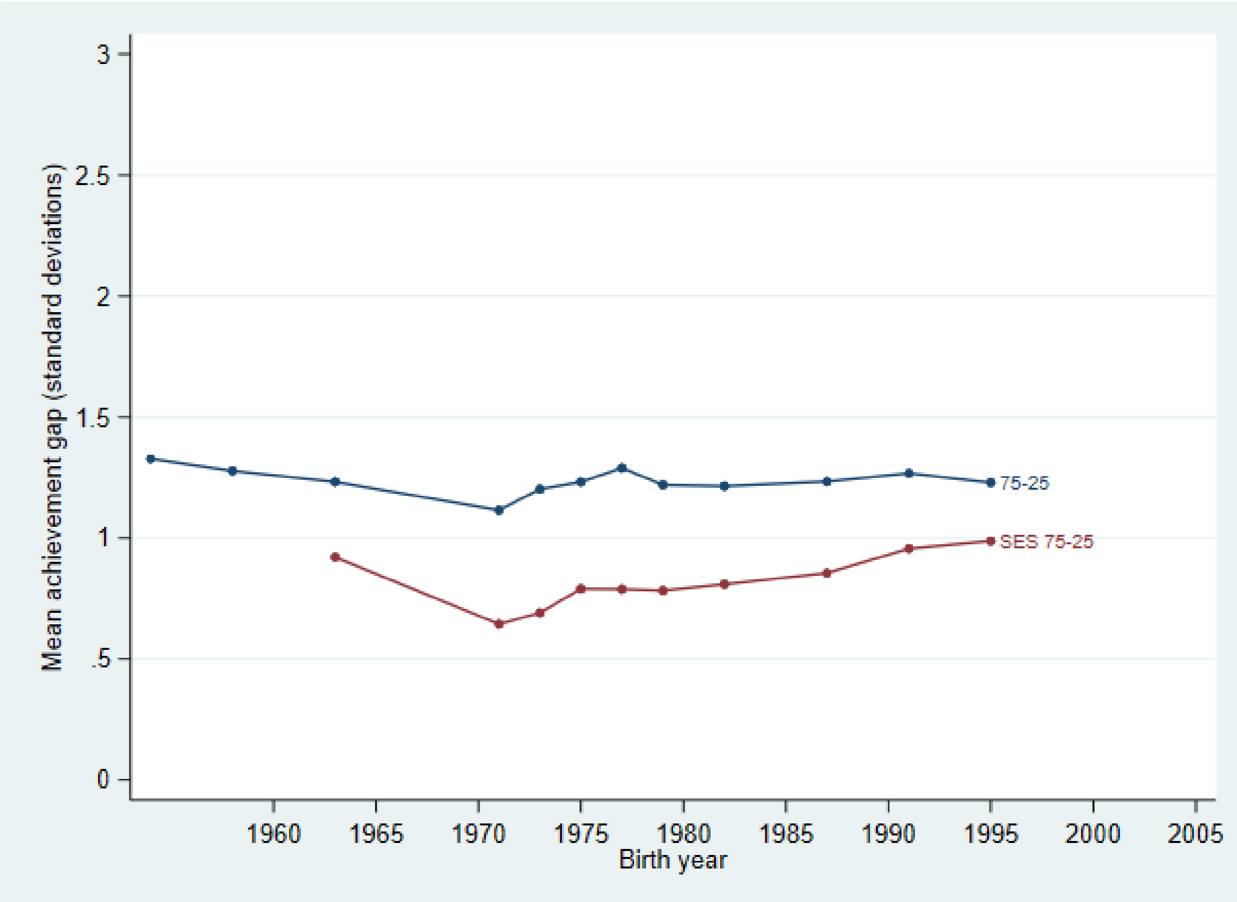


Figure c6. LTT-NAEP Unconditional and SES 75-25 Gaps, Age 17 Reading

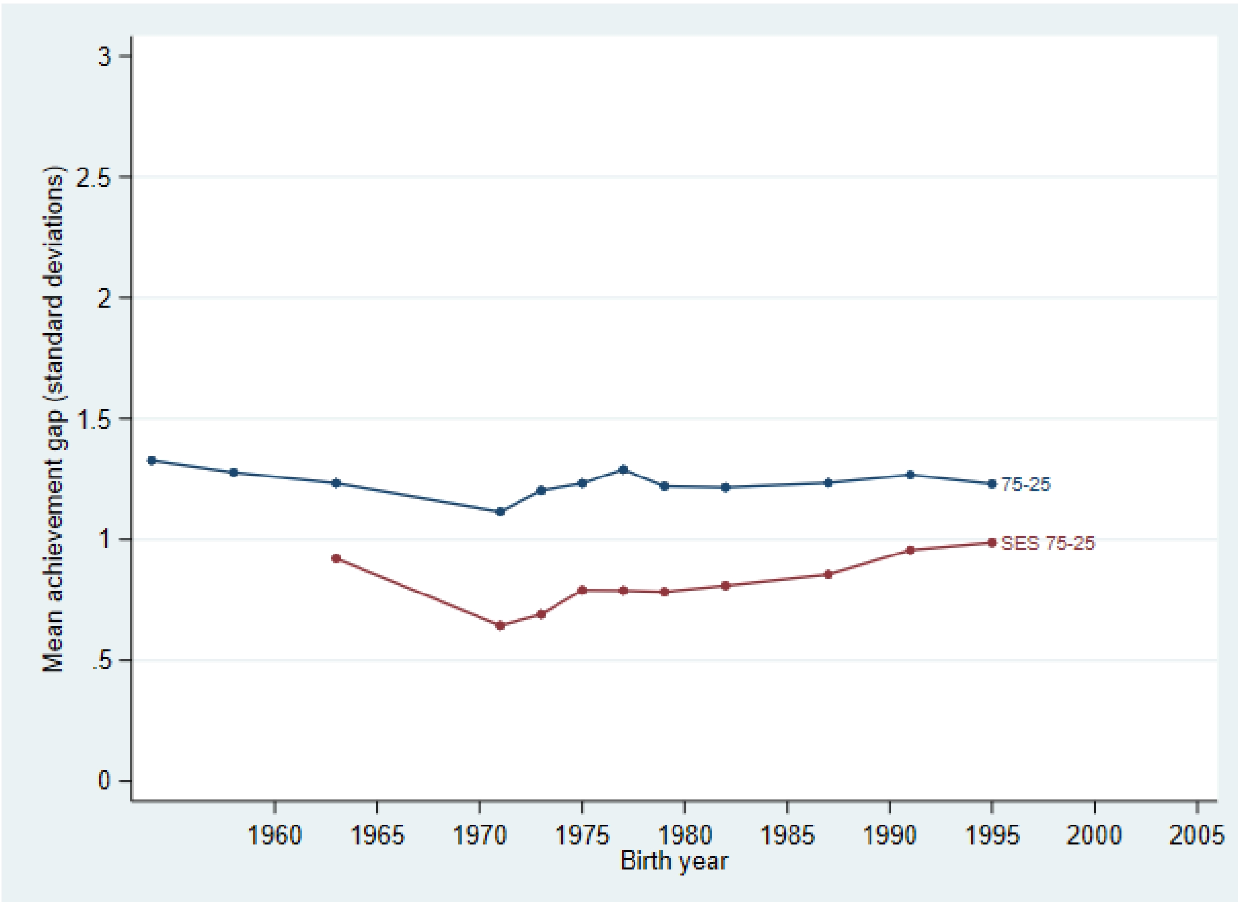


Figure c7. TIMSS Unconditional and SES 75-25 Gaps, Grade 8 Math

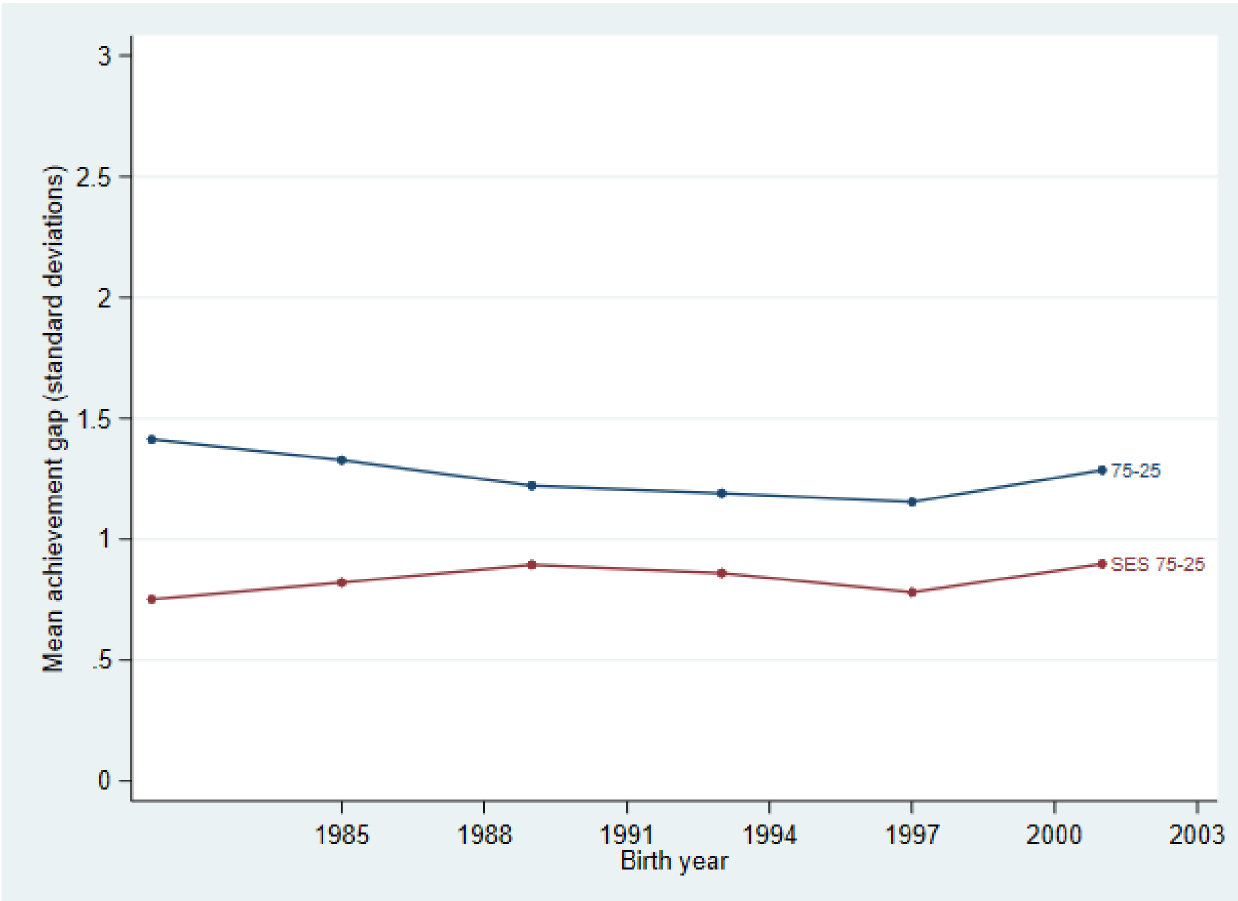


Figure c8. TIMSS Unconditional and SES 75-25 Gaps, Grade 8 Science

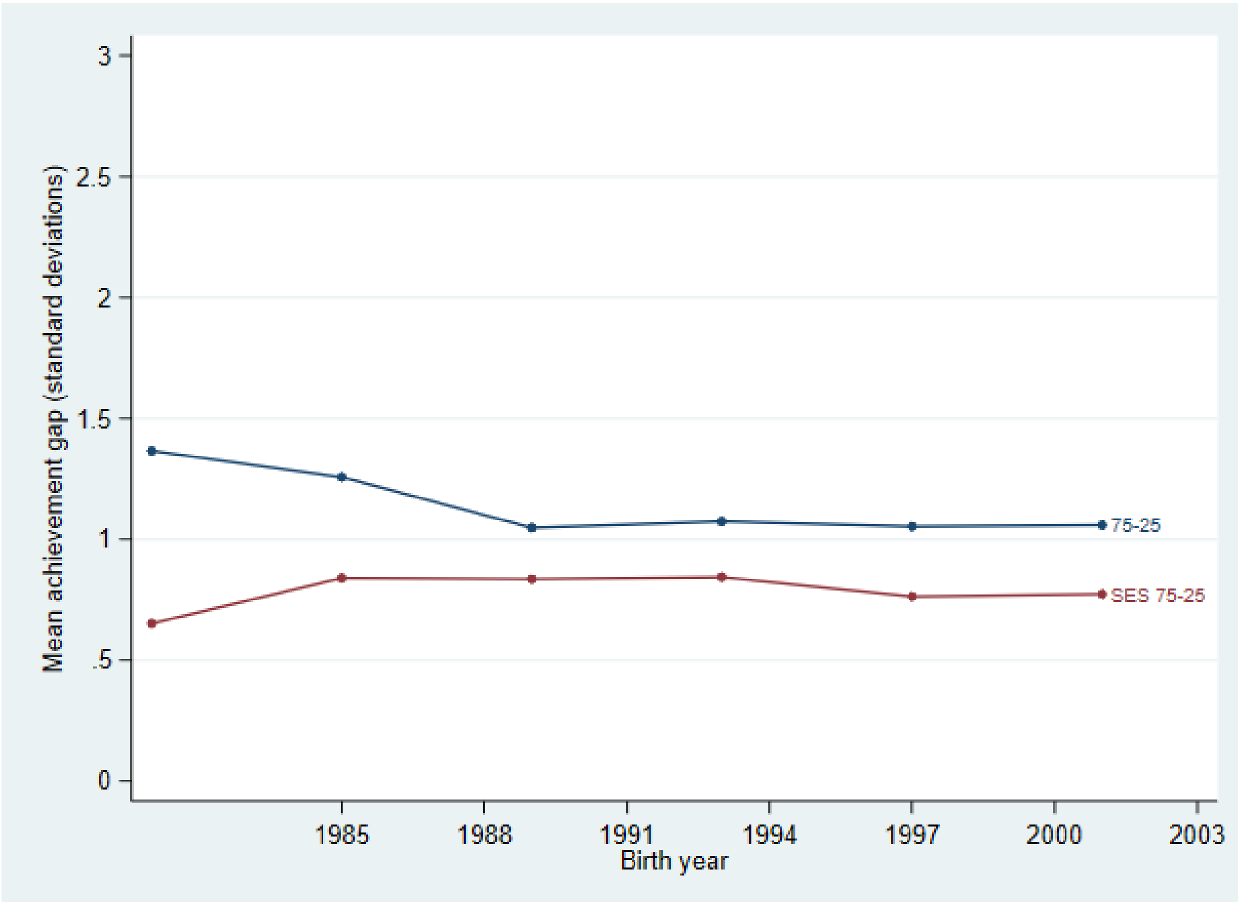


Figure c9. PISA Unconditional and SES 75-25 Gaps, Age 15 Math

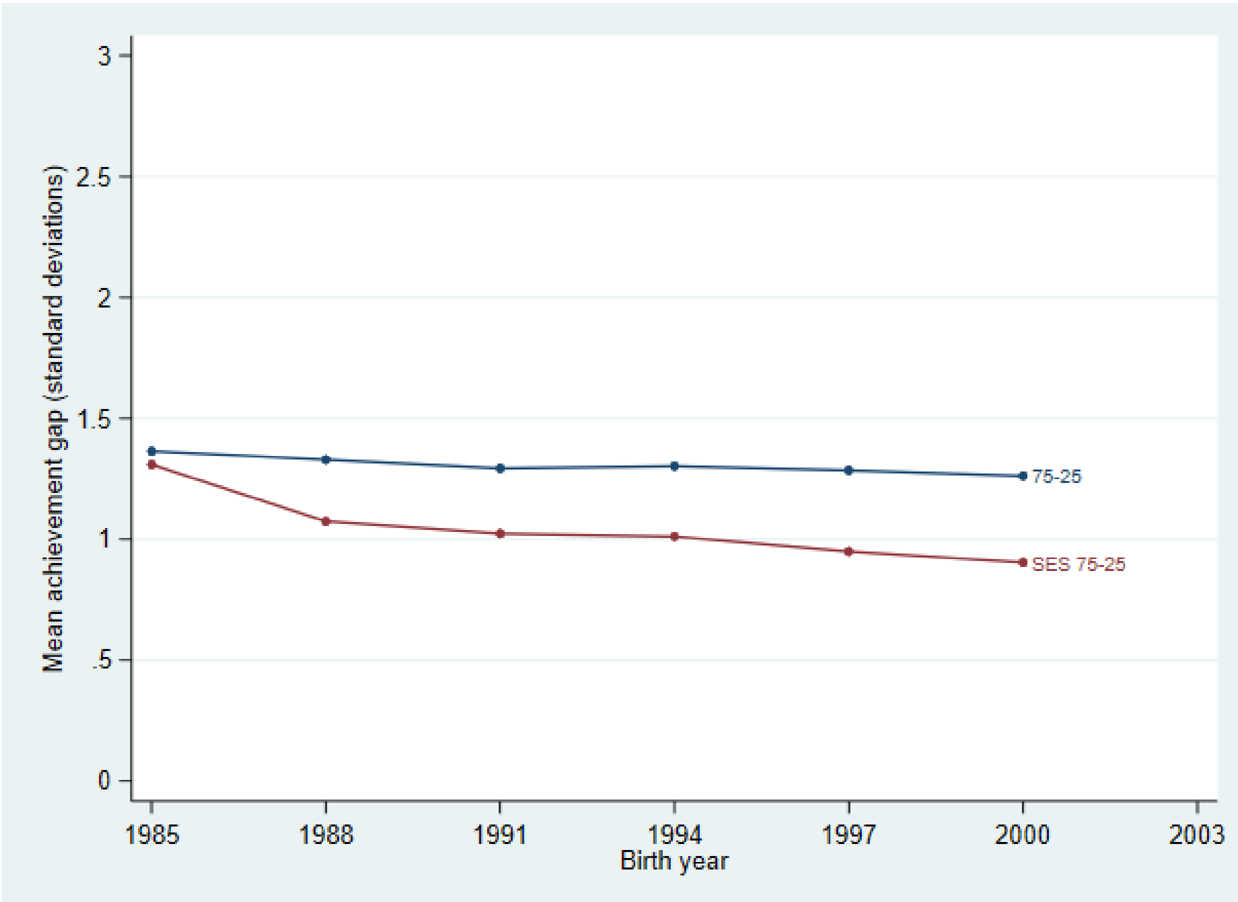


Figure c10. PISA Unconditional and SES 75-25 Gaps, Age 15 Reading

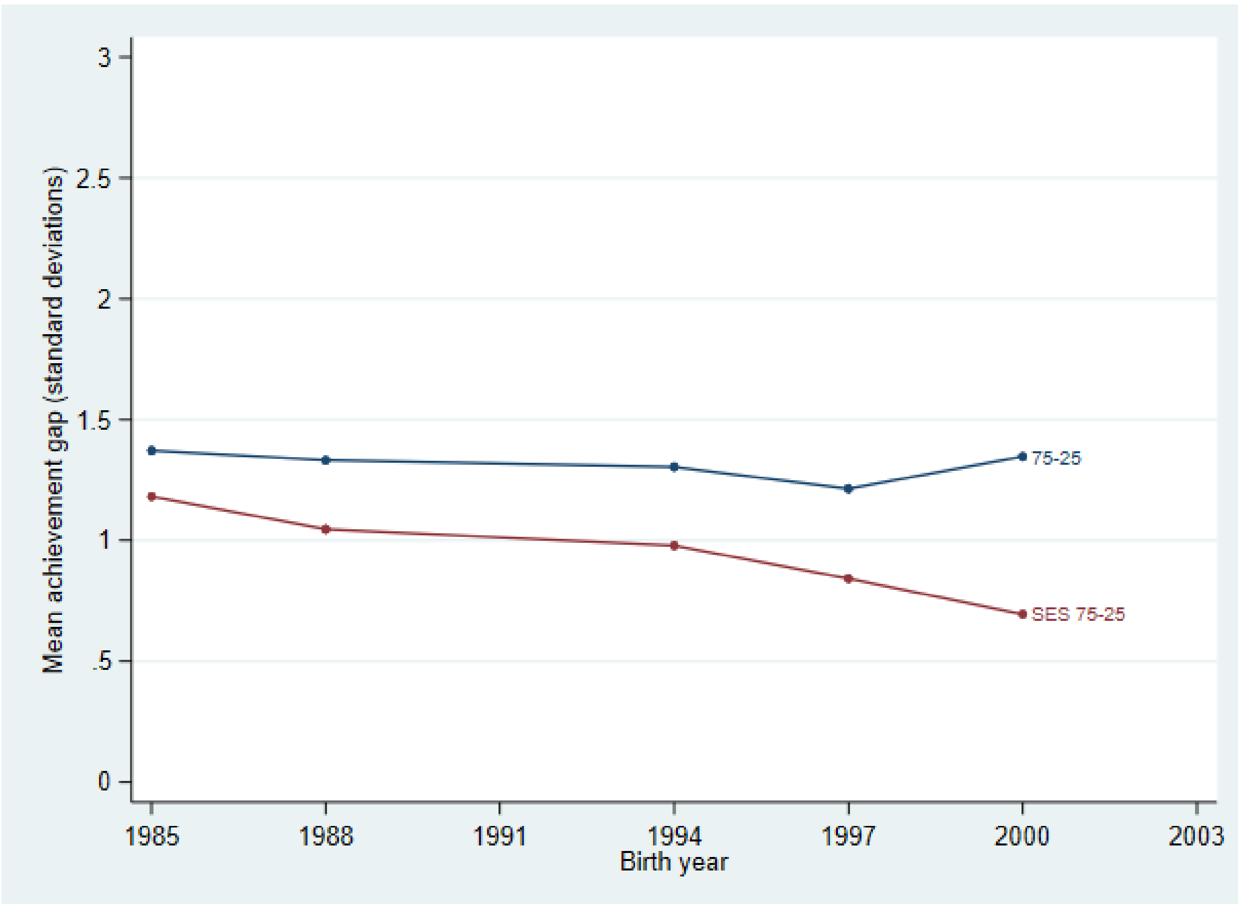


Figure c11. PISA Unconditional and SES 75-25 Gaps, Age 15 Science

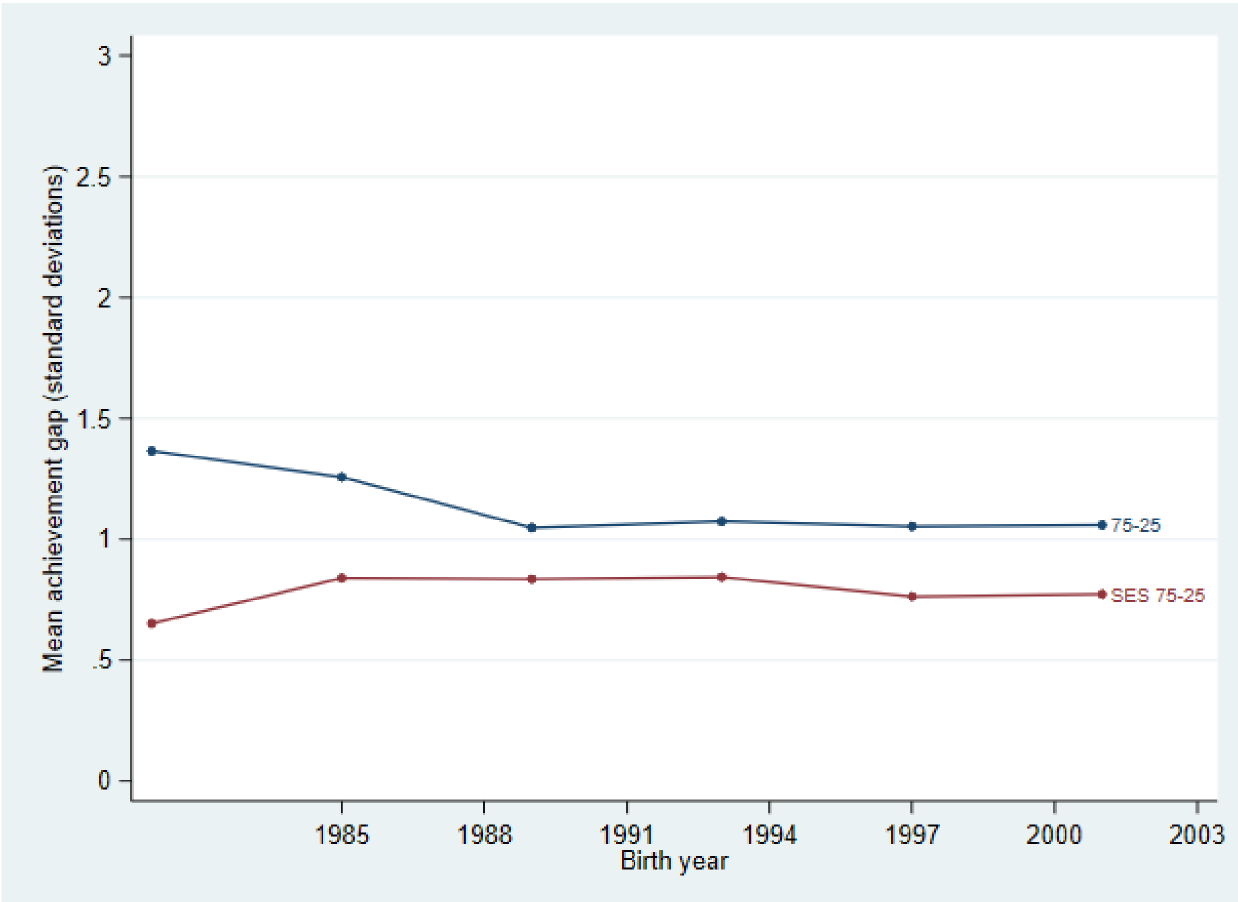


Table 1. Components of Background Surveys, NAEP 1990 and PISA 2015

NAEP 1990	PISA 2015
Parents' Education Level 1 = Didn't finish h.s. 2 = Graduated h.s. 3 = Some ed after h.s. 4 = Graduated college	**Highest education of parents** 0 = None 1 = Grade 6 2 = Grade 9 3 = <HS 4 = HS grad 5 = Associate's 6 = Bachelor's
Home possessions Do you have books in your home? Do you have magazines delivered regularly to your home? Do you have a newspaper delivered regularly to your home? Do you have an encyclopedia in your home?	**Home Possessions** Do you have your own room? Do you have educational software in your home? Do you have a link to the internet in your home? Do you have a dictionary in your home? Do you have a a quiet place to study in your home? Do you have a desk to study at in your home? Do you have textbooks in your home? Do you have classic literature in your home? (e.g., Shakespeare) Do you have books of poetry in your home? Do you have works of art in your home? (e.g., paintings) How many televisions do you have in your home? (None, one, two, three or more) How many computers do you have in your home? (None, one, two, three or more) How many musical instruments do you have in your home? (None, one, two, three or more) How many cars do you have at your home? (None, one, two, three or more) How many bathrooms do you have in your home? (None, one, two, three or more) How many books do you have in your home? (None, 1-10, 11-50, 51-100, 101-250, 251-500, more than 500) Do you have a computer you can use for school work in your home? Do you have a guestroom in your home? Do you have high-speed internet in your home? Do you have musical instruments in your home? Do you have technical reference books in your home? Do you have books on art, music, or design in your home?

Table 2: Description of Achievement Data

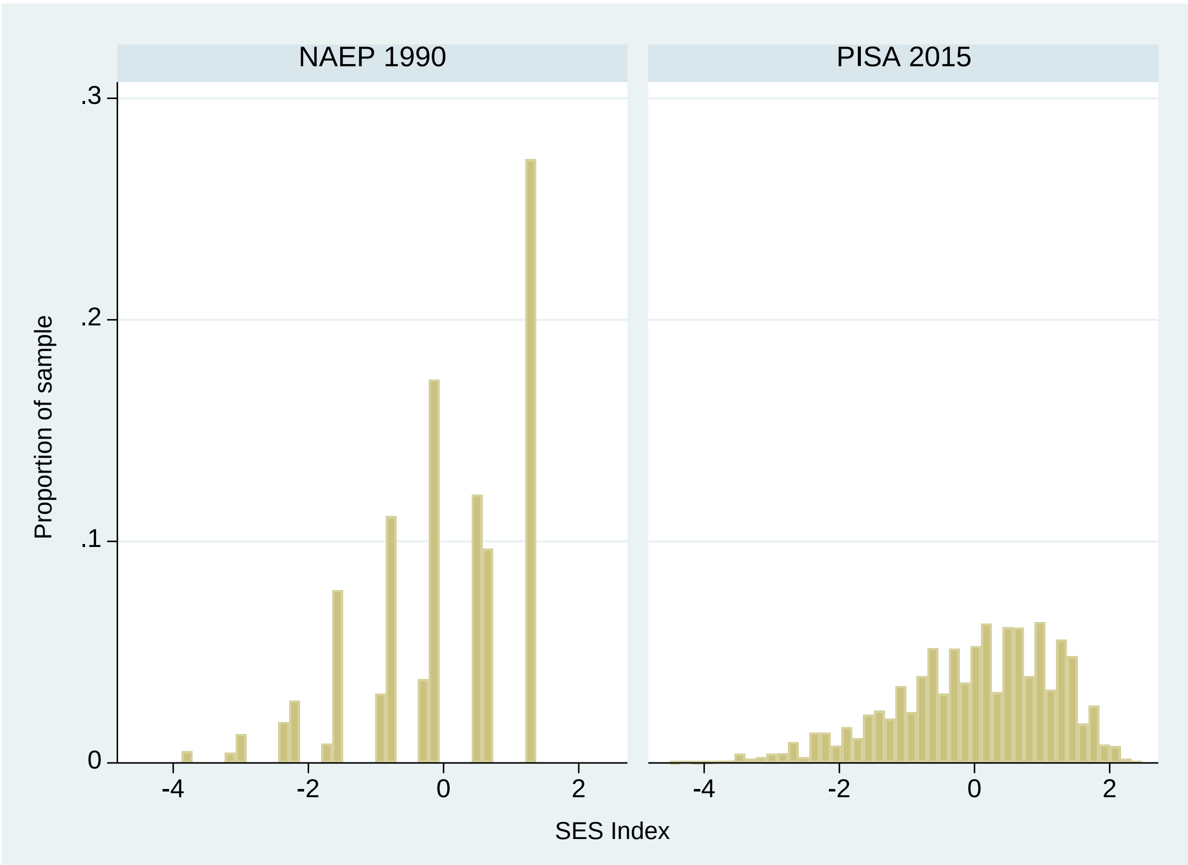
Test	Age/ grade	Birth cohorts		Observations by Test and Subject			
				Math	Reading	Science	Total
LTT-NAEP	age 13	1958-1999	Waves:	12	12	–	24
			Students:	99,450	115,780		215,230
LTT-NAEP	age 17	1954-1995	Waves:	12	12	–	24
			Students:	88,740	108,450		197,190
Main-NAEP	grade 8	1976-2001	Waves:	10	11	–	21
			Students:	1,004,650	1,122,980		2,127,630
TIMSS	grade 8	1981-2001	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.

Table 3. Racial SES Distribution over Time

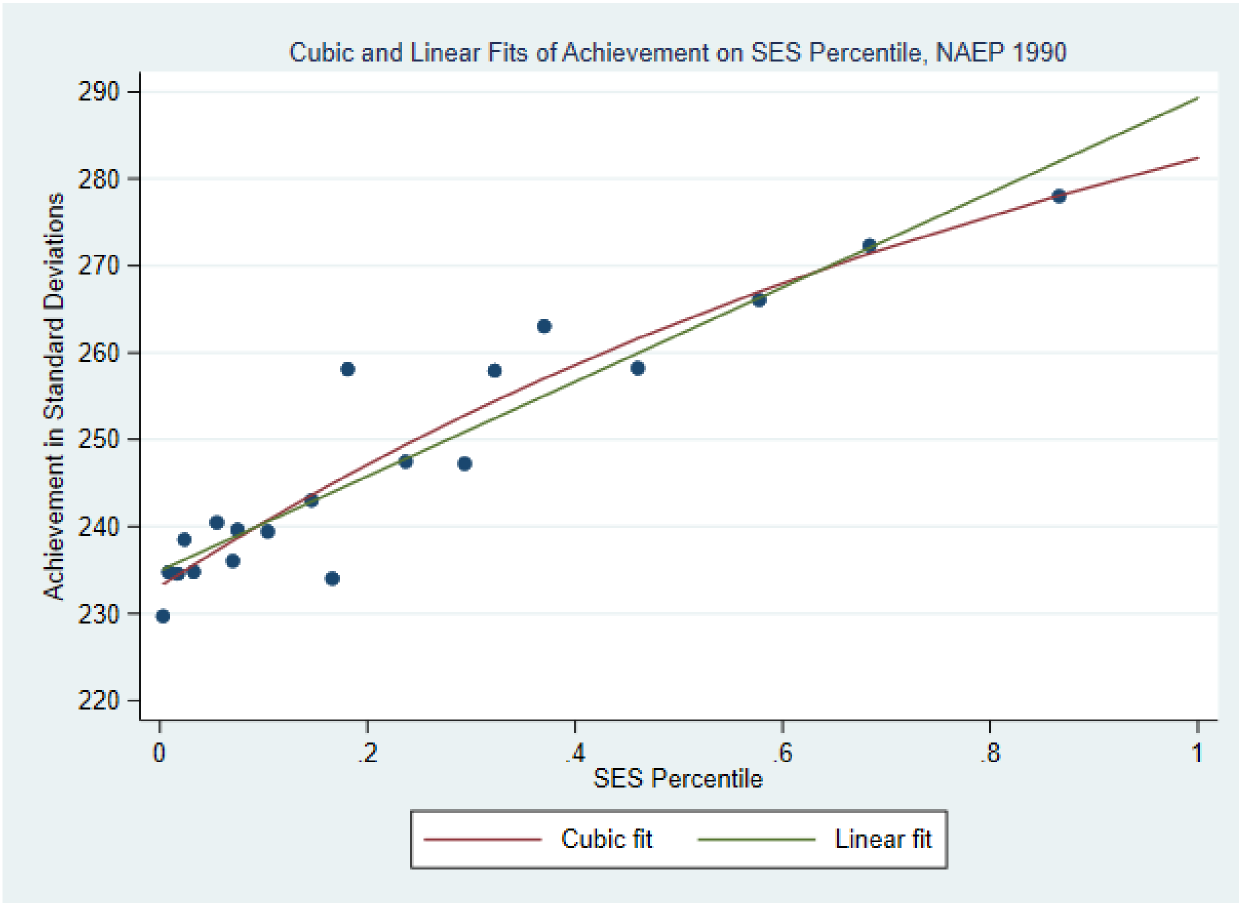
Test year	Birth year	% white	% black
Top 25% of aggregate SES			
1978	1961	27%	8%
2012	1995	33%	14%
Bottom 25% of aggregate SES			
1978	1961	20%	48%
2012	1995	14%	31%

Figure 1. Histogram of SES Distribution, NAEP 1990 and PISA 2015



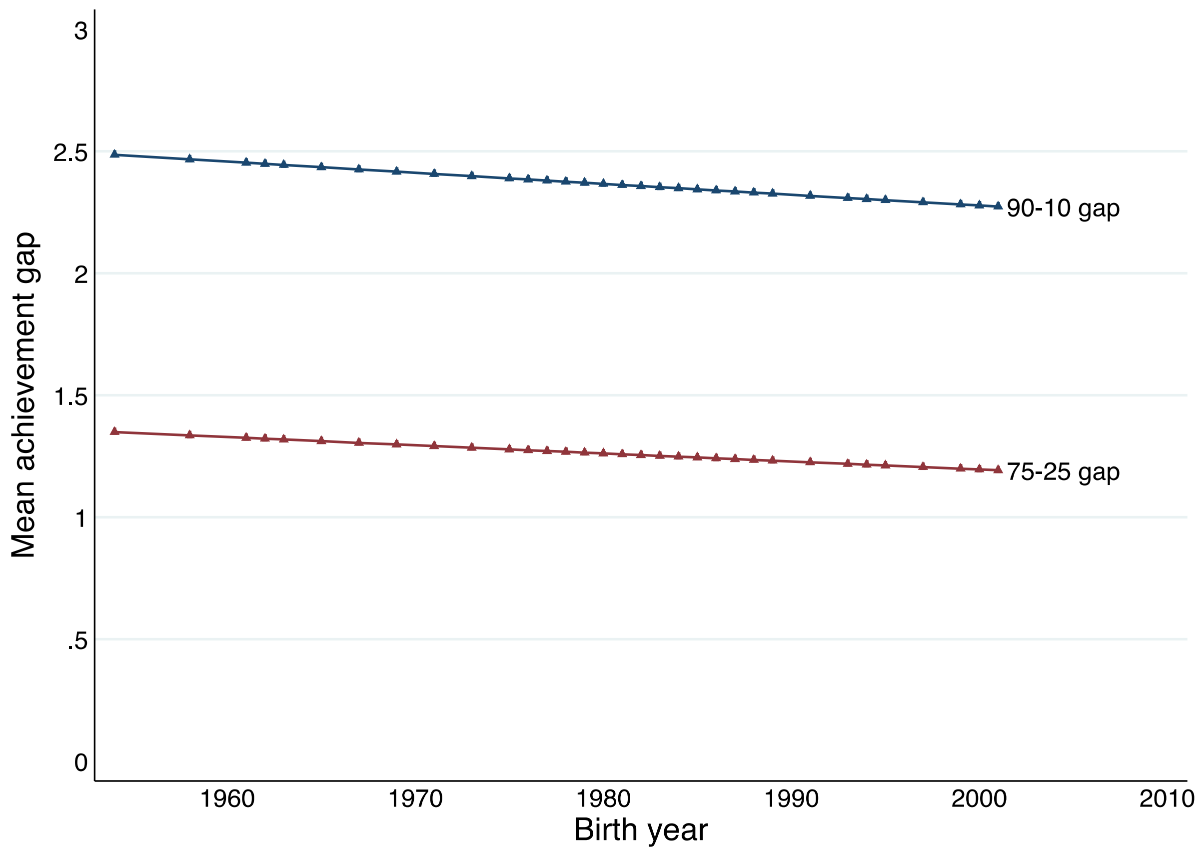
Notes: Frequency of specific values for the NAEP 1990 and PISA 2015 SES distributions calculated by the first principle component of the parent education and home possession variables; see text.

Figure 2. Linear and cubic regressions of NAEP 1990 scores on SES Percentiles



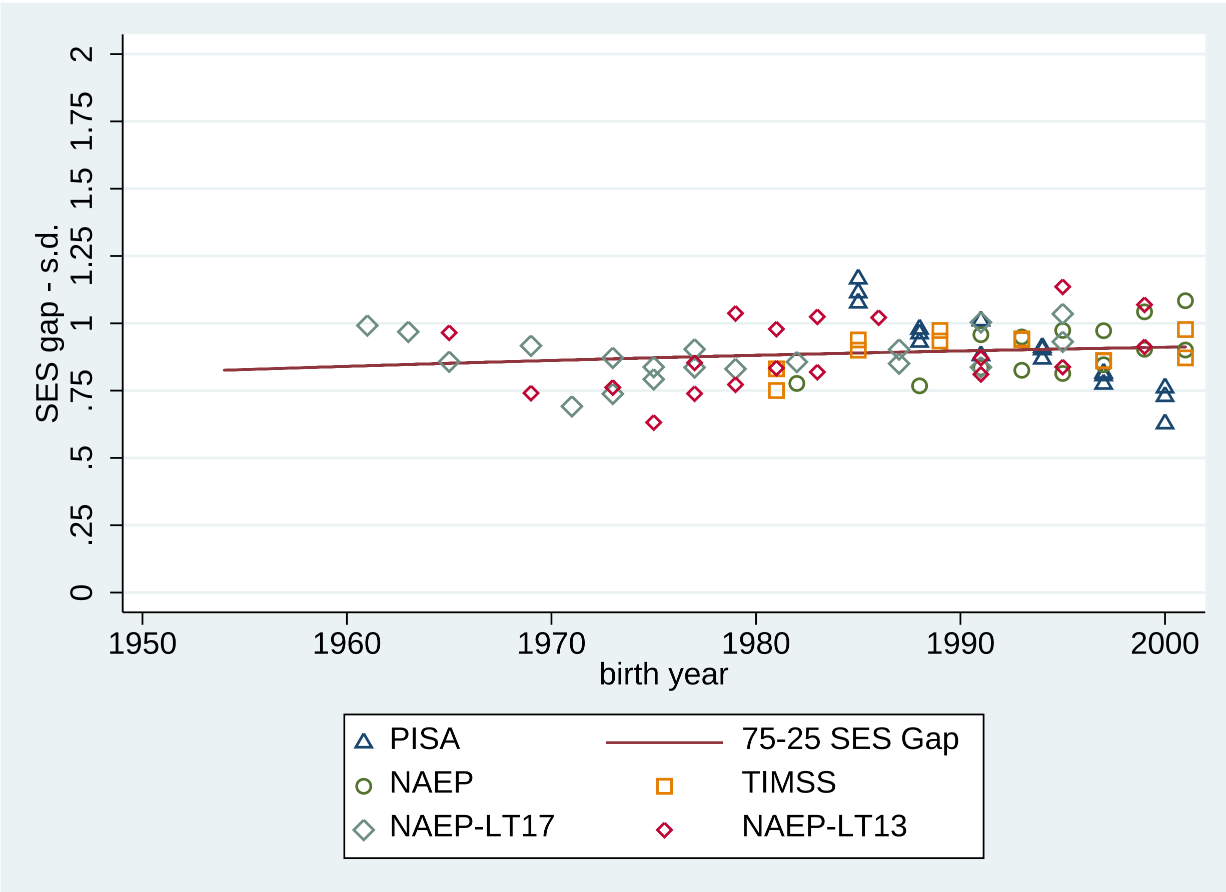
Notes: NAEP 1990 scores compared to midpoint of percentiles of the discrete values of the SES with linear and cubic relationships.

Figure 3: Unconditional Achievement Gaps of U.S. Students, Birth Cohorts 1954-2001



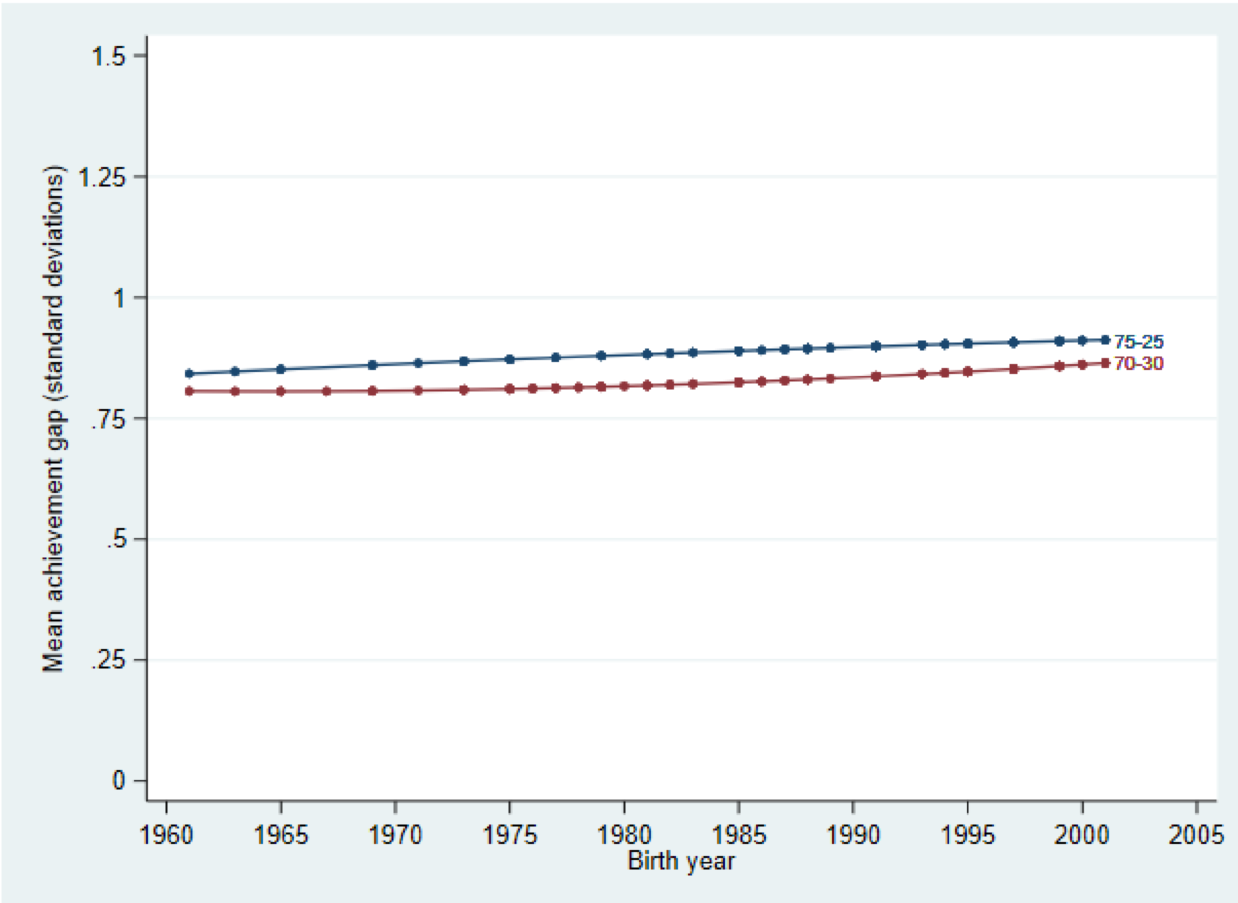
Note: All tests administered by LTT-NAEP, Main-NAEP, PISA, and TIMSS. 1954-2001 birth cohorts, all subjects, all students. Normalized achievement is measured in standard deviations. The s.d. presented is the difference between the year the test was administered and 2000 (or the closest year to that date available for a specific test series). Each marker indicates years where there are one or more underlying observations.

Figure 4. Trend in 75-25 SES Gap for 1961-2001 Birth Cohorts with Underlying Test Data



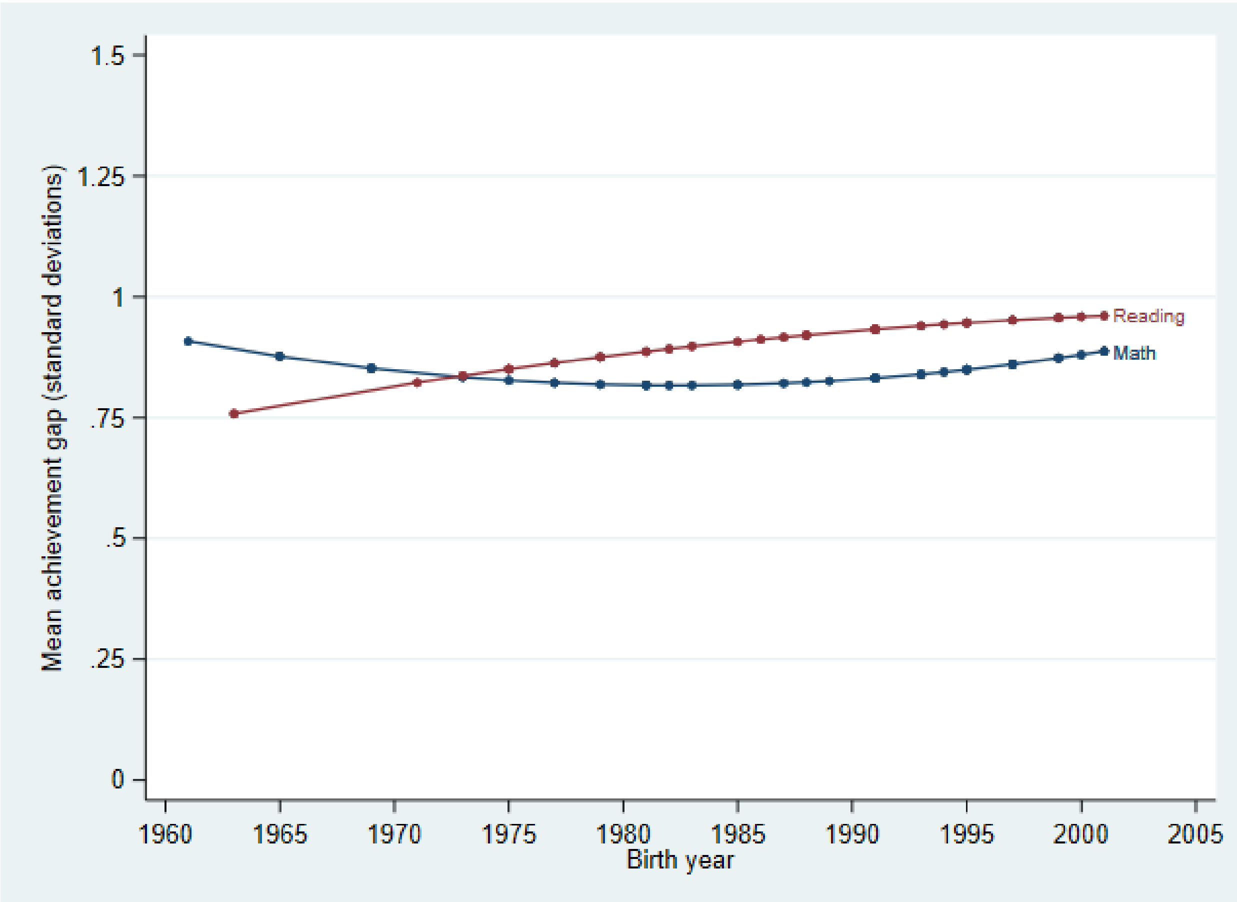
Notes: The line for the 75-25 SES Gap is the fitted quadratic from equation 2. Both the linear and quadratic terms are individually insignificantly different from zero ($|t| = 0.12$ for each); jointly, $F(2,72)=0.53$, $p>.59$. Test data points are adjusted by the fixed effects estimated for equation 2.

Figure 5. 75-25 and 70-30 SES Achievement Gaps, 1961-2001 Birth Cohorts



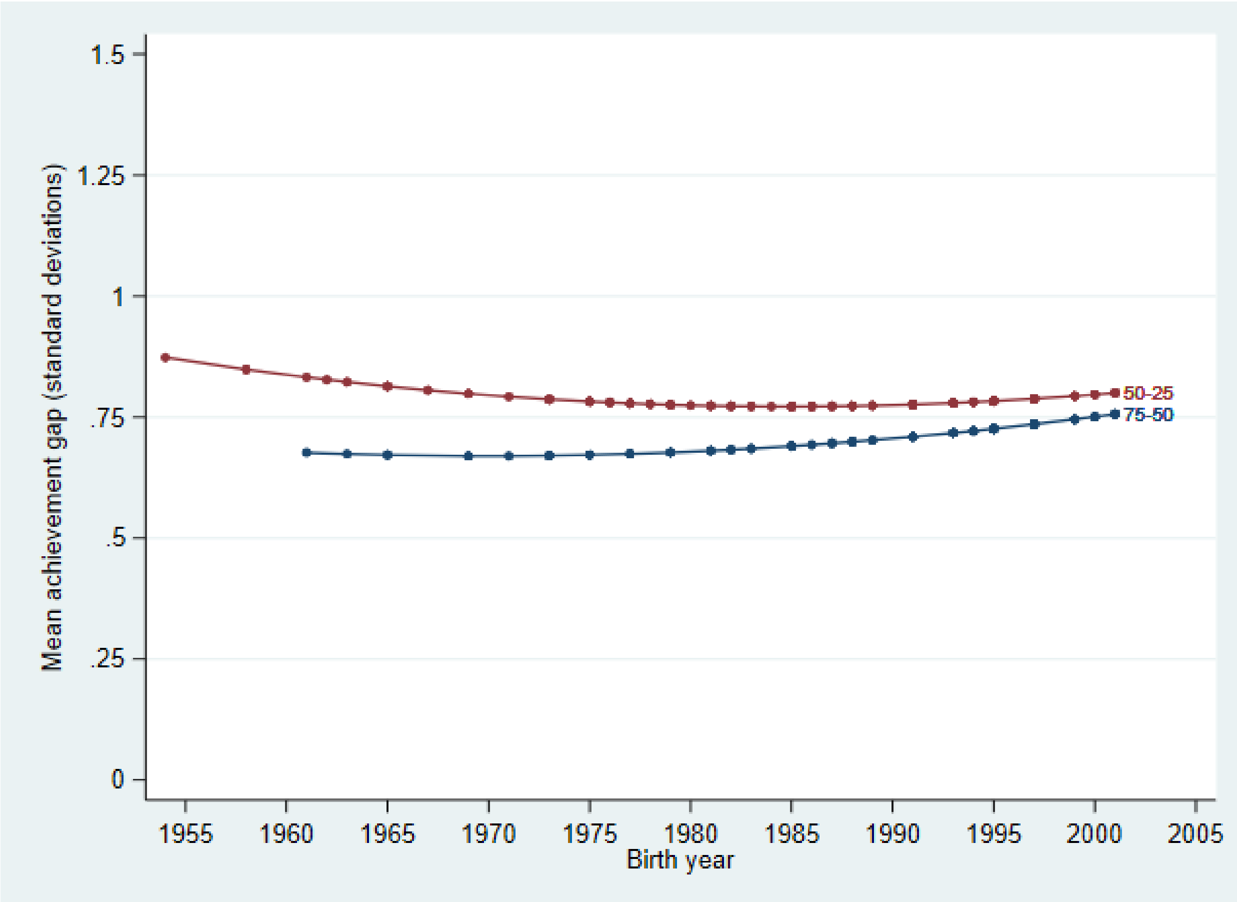
Notes: The markers on the trend lines indicate years where there is reliable information about the SES gaps. The 75-25 SES line is the same as shown in Figure 4. The 70-30 SES gap trend line is estimated from 91 observations; the joint significance test for the quadratic parameters is $F(2, 82)=1.05$, $p>0.35$.

Figure 6. 75-25 SES Achievement Gaps by Subject, 1961-2001 Birth Cohorts



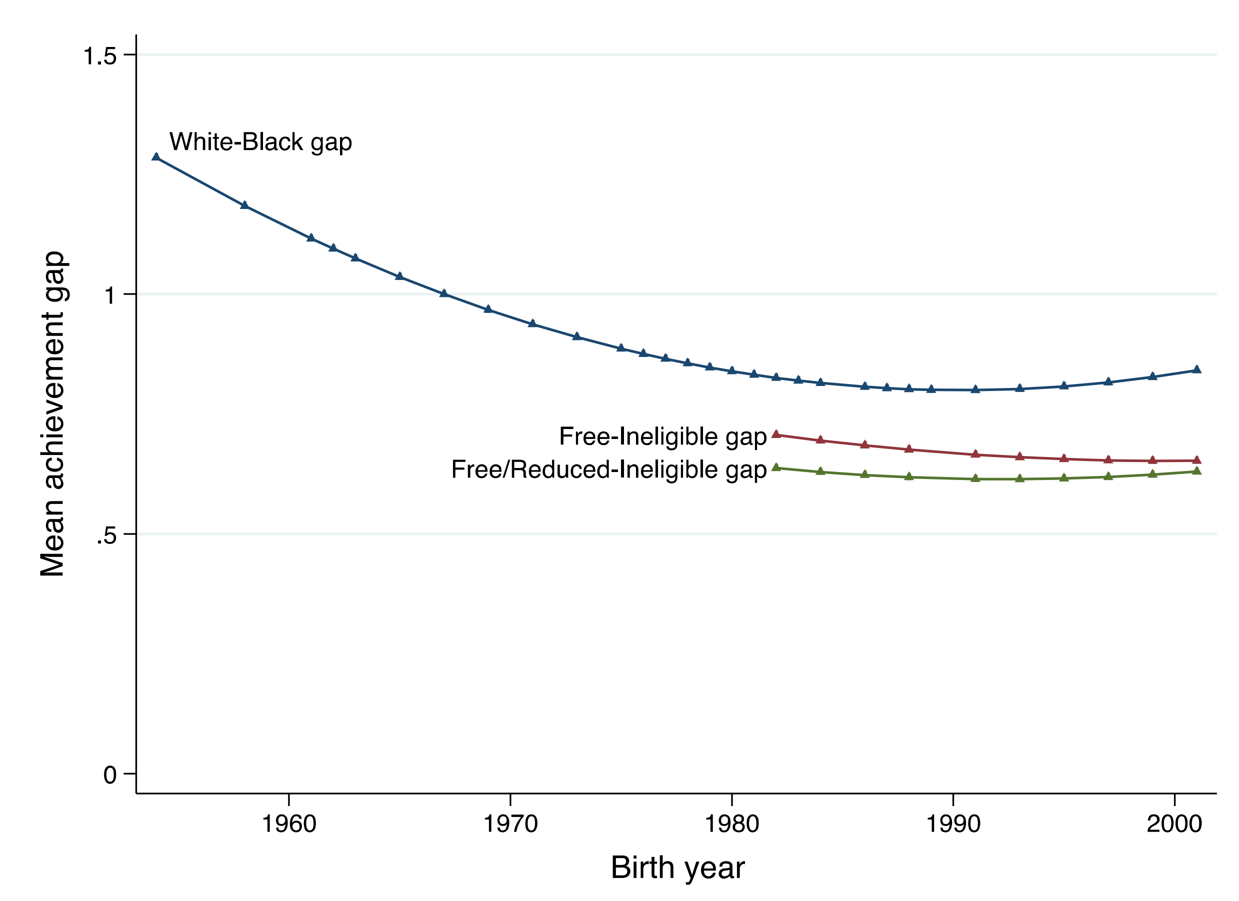
Note: See Figures 4 and 5 for data and methods.

Figure 7. 75-50 and 50-25 SES Achievement Gaps, 1954-2001 Birth Cohorts



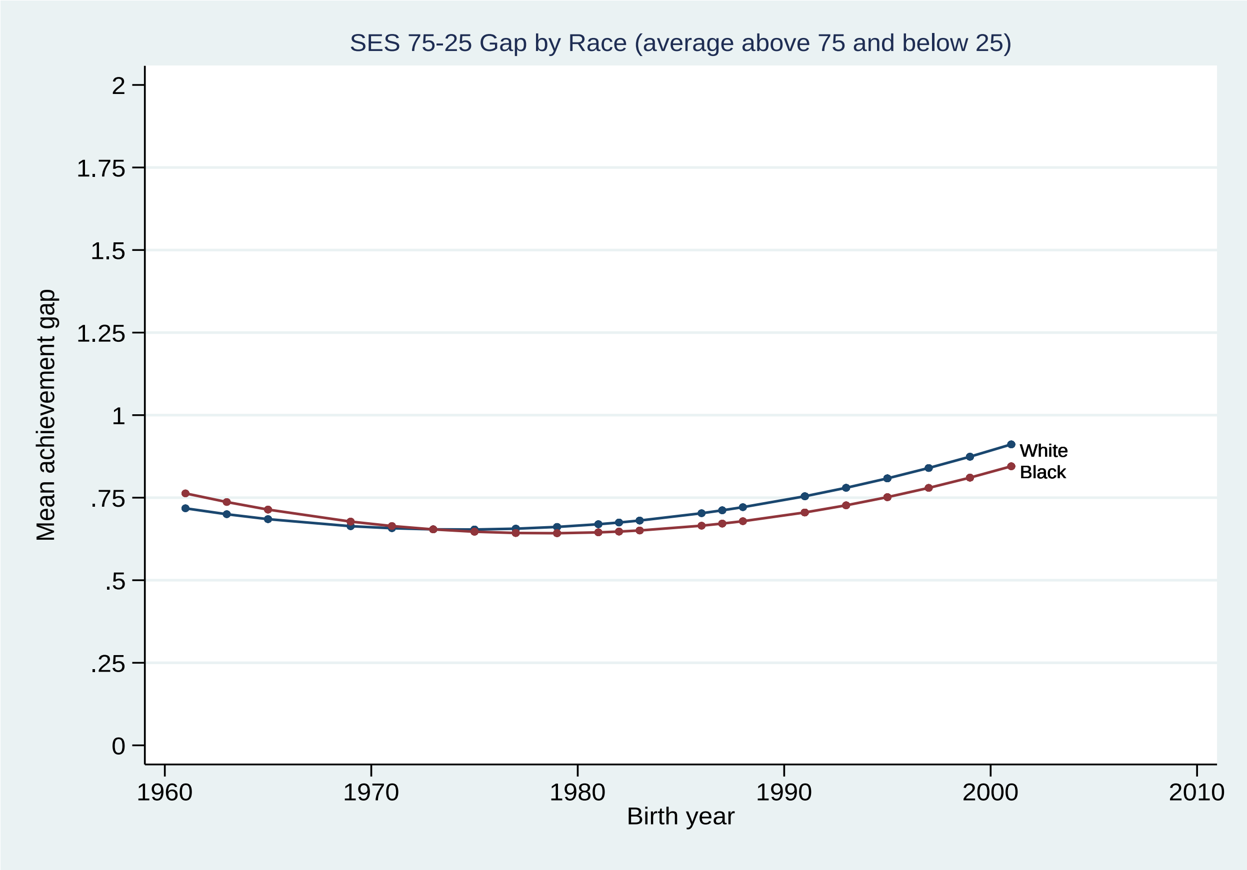
Note: See Figure 4 and 5 for data and methods.

Figure 8: Achievement Gaps for Eligibility for Free and Reduced-Price Lunch and for Race, Birth Cohorts 1954-2001



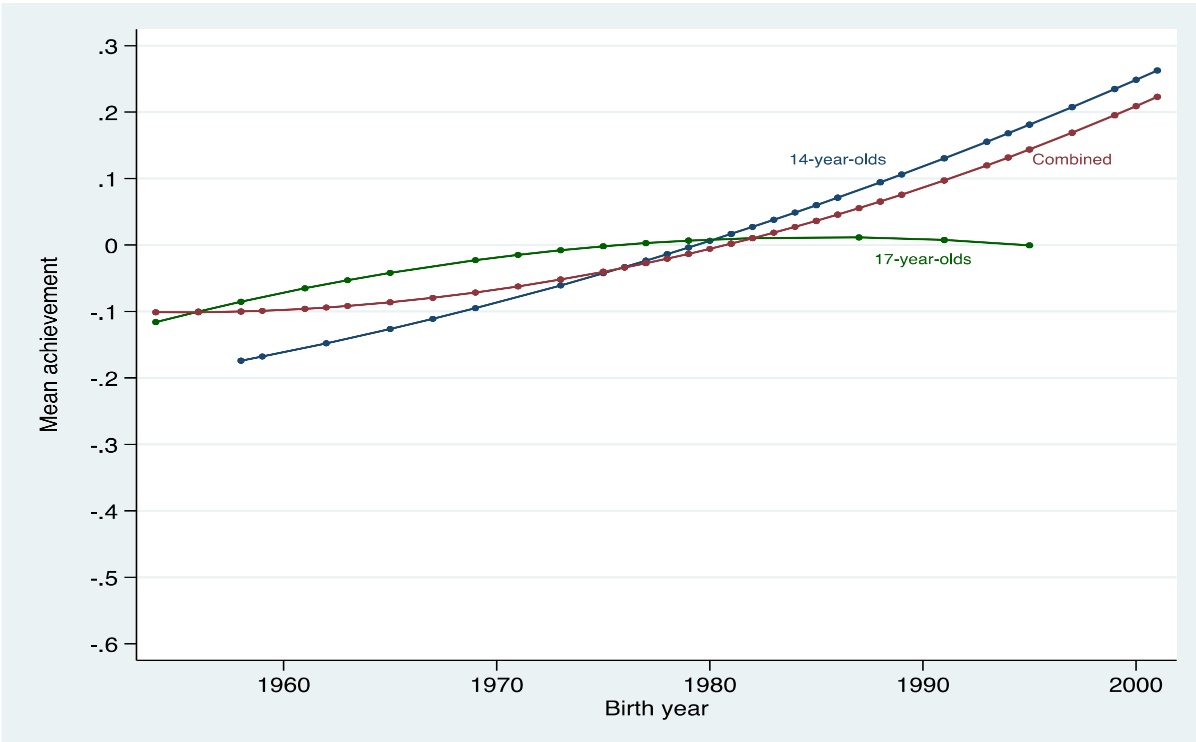
Note: Samples: For free and reduced price lunch, 1982-2001 birth cohorts, Main-NAEP surveys, math and reading, all students; for White-Black gap, 1954-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, black and white students. See Figure 4 for data and methods. Data on free and reduced-price lunch eligibility are only available for Main-NAEP tests, starting with the 1982 birth cohort.

Figure 9: 75-25 SES Achievement Gaps among White and Black Students Separately, Birth Cohorts 1961-2001



Note: Sample: 1954-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, white and black students. See Figure 4 for data and methods with national (common) SES distribution.

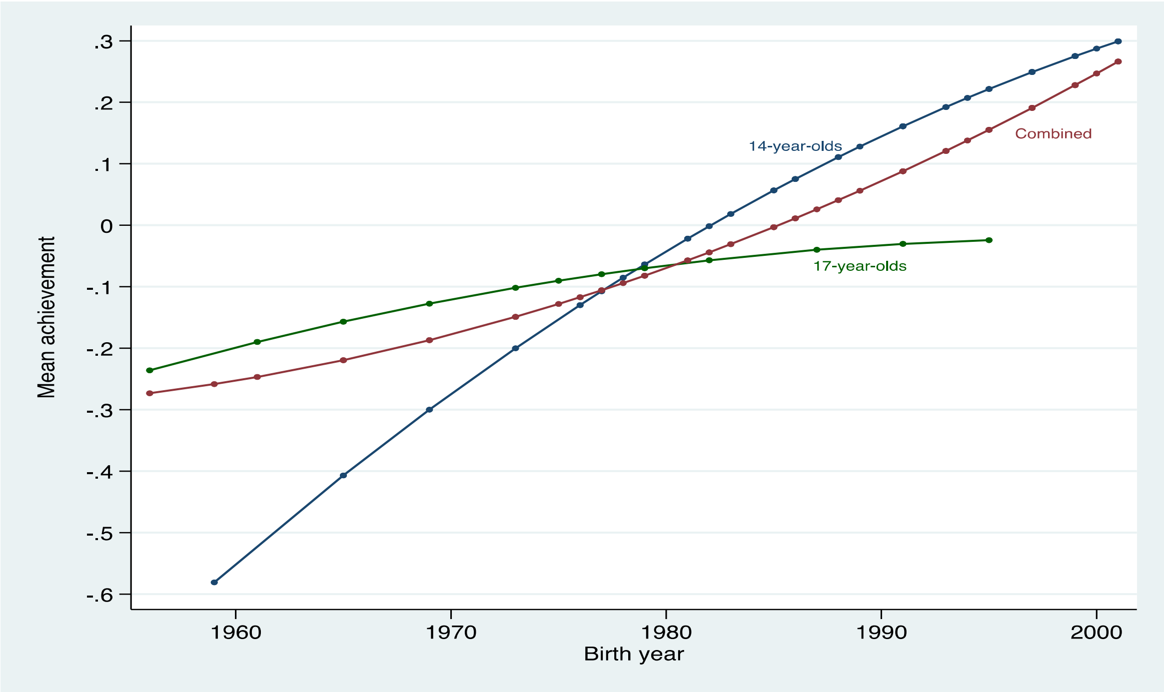
Figure 10: Achievement Levels of Younger and Older Students, Birth Cohorts 1954-2001



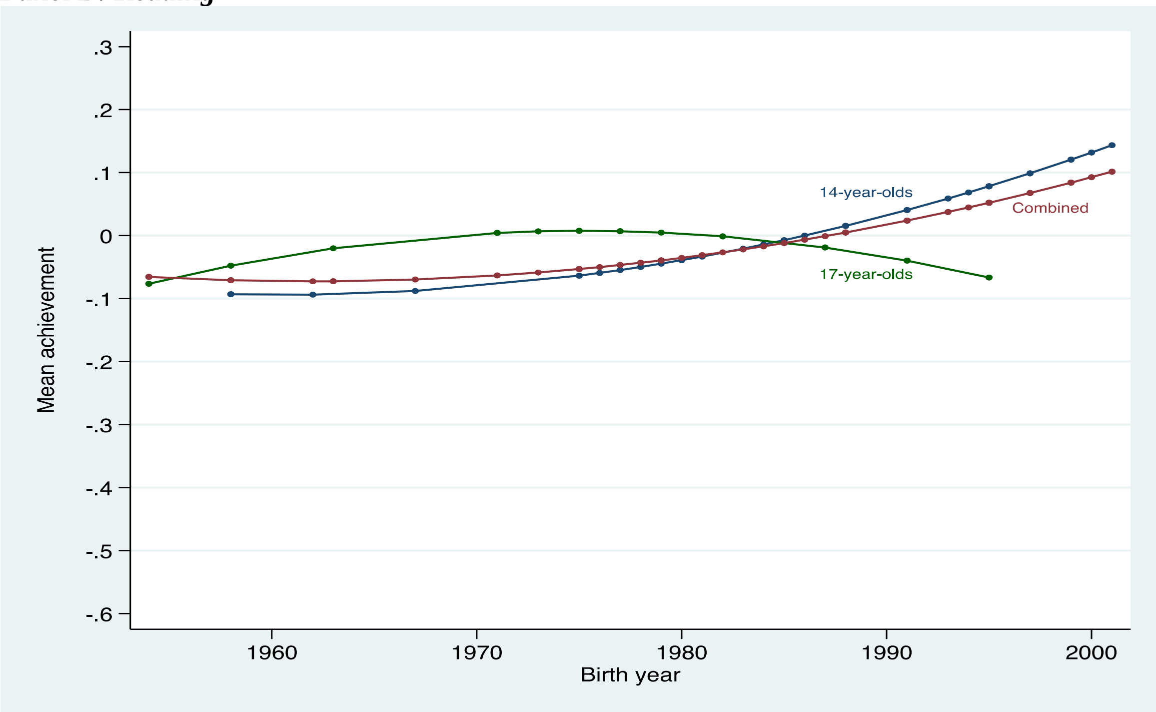
Note: Sample: 1954-2001 birth cohorts, all surveys, all subjects, all students. Younger students are those between ages 13 and 15 or in 8th grade, depending on the test. For expositional purposes, younger students are referred to as 14-year-olds. Older students are those aged 17 or in 12th grade, depending on the test. See Figure 4 for data and methods.

Figure 11: Achievement Levels of Younger and Older Students by Subject, Birth Cohorts 1954-2001

Panel A: Math

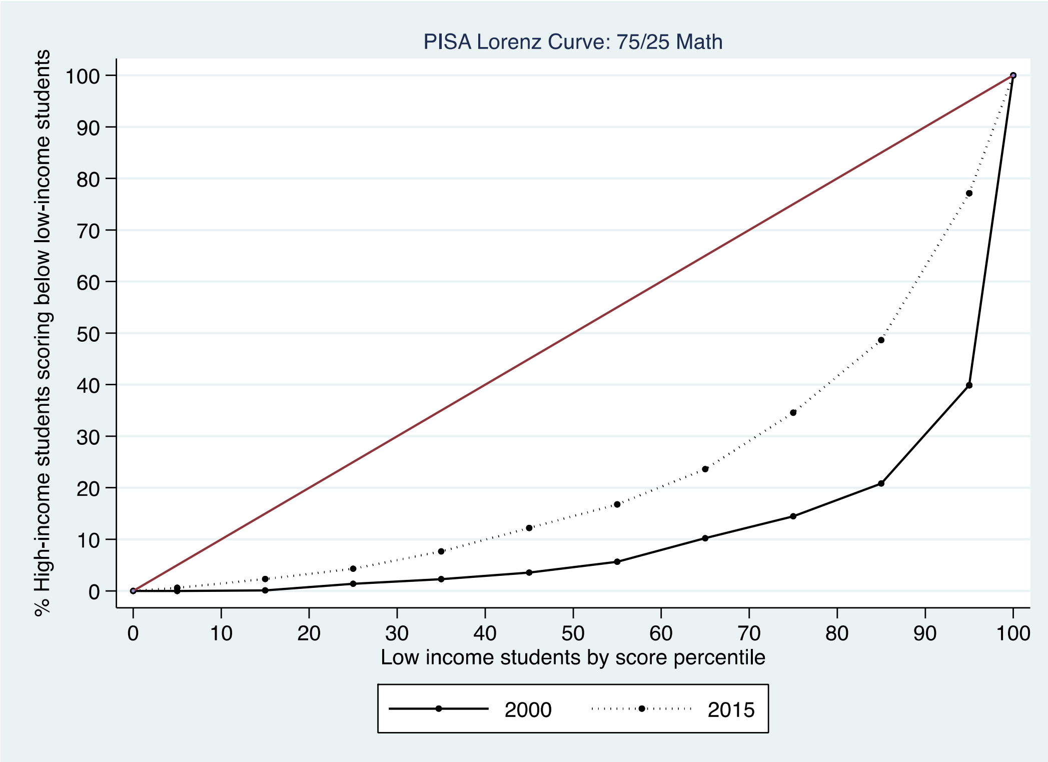


Panel B: Reading



Note: See Figures 4 and 5 for data and methods.

Figure 12. Lorenz Curves for 75-25 and 50-25 SES Achievement Distributions, PISA 2000 and PISA 2015



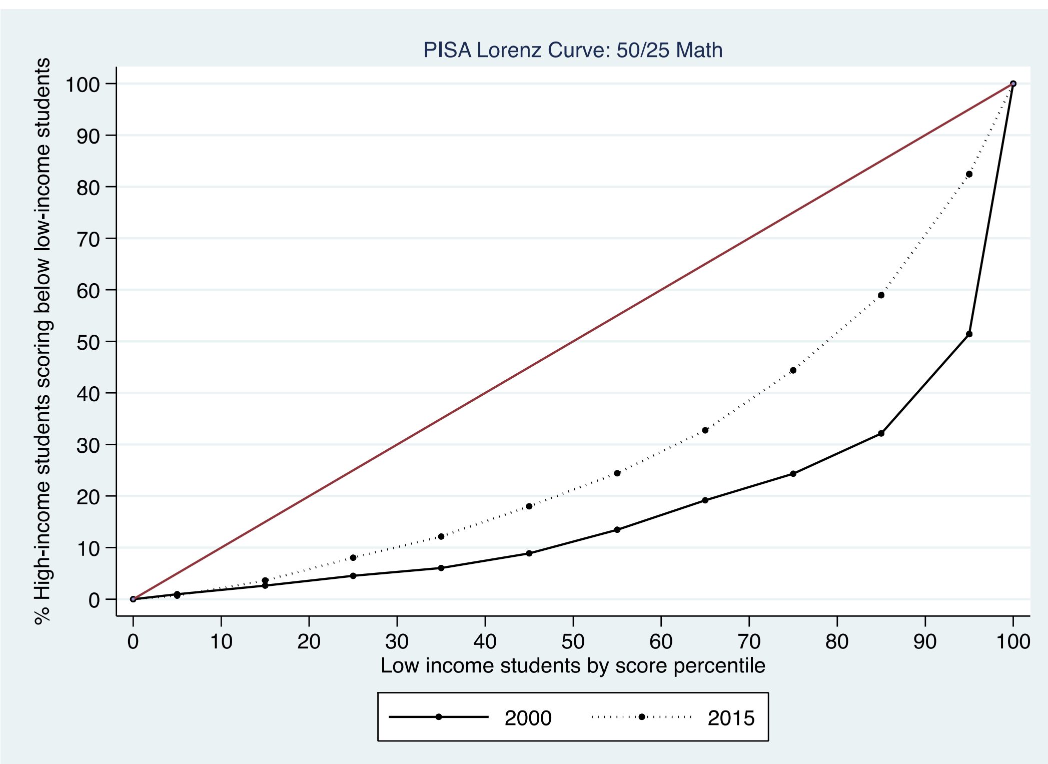


Figure 13. Lorenz Curves for 75-25 and 50-25 SES Achievement Distributions, TIMSS 1995 and TIMSS 2015