ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hanushek, Eric A.; Peterson, Paul E.; Talpey, Laura M.; Woessmann, Ludger

Working Paper

Long-Run Trends in the U.S. SES-Achievement Gap

IZA Discussion Papers, No. 12971

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Hanushek, Eric A.; Peterson, Paul E.; Talpey, Laura M.; Woessmann, Ludger (2020) : Long-Run Trends in the U.S. SES-Achievement Gap, IZA Discussion Papers, No. 12971, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/215367

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12971

Long-Run Trends in the U.S. SES-Achievement Gap

Eric A. Hanushek
Paul E. Peterson
Laura M. Talpey
Ludger Woessmann

FEBRUARY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12971

Long-Run Trends in the U.S. SES-Achievement Gap

Eric A. Hanushek
Hoover Institution/Stanford University, NBER, CESifo and IZA

Paul E. Peterson
Harvard University and Hoover Institution/ Stanford University

Laura M. Talpey
Hoover Institution/Stanford University

Ludger Woessmann
University of Munich, ifo Institute, CESifo and IZA

FEBRUARY 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Long-Run Trends in the U.S. SES-Achievement Gap*

Rising inequality in the United States has raised concerns about potentially widening gaps in educational achievement by socio-economic status (SES). Using assessments from LTT-NAEP, Main-NAEP, TIMSS, and PISA that are psychometrically linked over time, we trace trends in achievement for 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 improved achievement levels, which have risen at age 14 but have remained unchanged at age 17 for the past quarter century.

JEL Classification: H4, I24, J24

Keywords: student achievement, inequality, socio-economic status, United States, NAEP, TIMSS, PISA

Corresponding author:
Ludger Woessmann
Ifo Institute for Economic Research at the University of Munich
Poschingerstr. 5
81679 Munich
Germany

E-mail: woessmann@ifo.de

* Helpful comments were received from Markus Broer, Greg Duncan, Glenn Ellison, John Klopfer, Magne Mogstad, Richard Murnane, Sean Reardon, Danish Shakeel, Abhijeet Singh, Chris Taber, and participants at the annual meetings of the American Economic Association and the Association for Public Policy Analysis and Management, the IRP Summer Research Workshop at the University of Wisconsin-Madison, the CESifo Economics of Education conference in Munich, and the Hoover Economic Policy Lunch at Stanford. This is a revised and refined version of Hanushek, Peterson, Talpey, and Woessmann (2019).

1. Introduction

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] To prevent poverty, Congress and many states enacted new education programs designed to enhance the human capital of children born into poor and otherwise disadvantaged households. In this paper, for the cohorts born from 1961 to 2001 we provide consistent evidence on changes in the gap in educational achievement between children raised within families of high and low socio-economic status (SES), as measured by student performance on standardized tests. Our main finding is that, despite the policy undertakings, the SES-achievement gap remains as large today as it was when the poverty war was declared. To the extent that these tests are predictive of a child's future prosperity, our results imply that intergenerational mobility is unlikely to improve in the middle decades of the 21st Century.

In advanced industrial societies, cognitive skills are highly correlated with economic outcomes. Indeed, the U.S. labor market rewards cognitive skills more than almost all other developed countries (Hanushek, Schwerdt, Wiederhold, and Woessmann (2015, 2017)). It is thus for good economic reasons that President Johnson and others have long searched for tools that could break the linkage between SES and student learning (Ladd (1996); Carneiro and Heckman (2003); Krueger (2003); Magnuson and Waldfogel (2008)).

Given the topic's importance, it is surprising that trends in SES-achievement gaps are 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."

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

The empirical basis for these conclusions is limited. In the seminal study of trends in SES-achievement gaps, Reardon (2011b) finds widening achievement gaps by SES measured by parental income. His innovative study combines cross-sectional data from twelve independent surveys to map the time-series pattern of achievement gaps. Unfortunately, that study suffers from severe measurement issues arising from data collection procedures that were plagued by missing and incomplete data and that relied on outmoded procedures and unusual survey choices. We show that, once measurement errors are taken into account, there is no evidence for widening SES-achievement gaps in his data.

We add to this sparse literature by providing the first comprehensive analysis of long-run trends in SES-achievement gaps from psychometrically linked data sets. We draw upon data from four well-documented surveys that have employed psychometric methods to link achievement in math, reading, and science over time: the Long-Term Trend assessment administered by the National Assessment of Educational Progress (LTT-NAEP), the Main-NAEP, the Trends in International Mathematics and Science Study (TIMSS), and the Programme for International Student Assessment (PISA). These assessments were administered to representative samples of U.S. adolescent students who were born over the nearly five decades between 1954 and 2001. Each test was designed to be comparable over time.

Using individual data for over two million students, we construct an index of SES based on information about parental education and home possessions of the students for 96 separate test-subject-age-year observations. This SES index allows us to measure SES-achievement gaps, with our main analysis focusing on the achievement difference between students in the top and bottom quartiles of the SES distribution on 81 testing occasions. Taking into account fixed effects for testing regimes, subjects, and age, we estimate a quadratic trend of the aggregate pattern in the SES-achievement gap over time.

Contrary to prior research and popular commentary, we find little change in the SES-achievement relationship across the past close to fifty years. The SES-achievement gap between the top and bottom SES quartiles (75-25 SES gap) has remained essentially flat at roughly 0.9 standard deviations (s.d.), a gap roughly equivalent to a difference of three years of learning between the average student in the top and bottom quartiles of the distribution. Moreover, we do not see much change in gaps at other points in the SES distribution. In none of the four sets of assessments do we observe the pronounced increase in the gaps identified in Reardon (2011b).

On the contrary, the trend tracked by PISA shows a decline over the most recent time period, while gaps observed in the other three assessments show either no change or only minor drifts upward. When data from all assessments are combined, the overall trend hardly wavers.[2]

These gaps remain steady within the context of quite stagnant levels of average achievement 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 middle-school students, those gains are no longer evident by age 17, the point students are expected to be ready for college and careers.

In robustness analyses, we show that our results are unaffected by consideration of a range of methodological issues. The main finding of a flat SES-achievement gap is confirmed in analyses of the achievement gap by subsidized lunch eligibility and in separate estimations by ethnicity that consider changes in the ethnic composition. Further robustness analyses include a treatment of limited information on the tails in the categorical SES data, an alternative point estimation approach to our baseline group calculation approach, and an alternative analysis that considers the ordinal nature of the underlying achievement data.

The next section reviews the literature on SES-achievement gaps. Section 3 describes our achievement data, and section 4 discusses our methodological approach. Section 5 reports our baseline evidence on trends in student achievement gaps and levels. Section 6 discusses various issues associated with the measurement of SES and provides supplementary analyses as robustness checks. The final section discusses and concludes.

2. Existing Literature on the SES-Achievement Gap

Definitions and measurement 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)). In this paper, we are interested in changes over time in the connection between SES and student achievement, a probable predictor of future economic opportunities. In that sense, it resembles studies of intergenerational mobility, which compare parental SES and the child's SES as an adult. After briefly

[2] We also show that the overall achievement dispersion in the population, while narrowing slightly, has shown limited change, and we confirm prior analysis that the black-white achievement gap, while strongly declining over the first half of the observation period, has stalled over the past quarter century.

summarizing research describing the relationship between indicators of family SES and student achievement, we look in more detail at research that traces temporal changes in this relationship.

2.1 The SES-Achievement Connection

The strong relationship between SES and achievement has long been known (Neff (1938)). Coleman et al. (1966), in their seminal study of *Equality of Educational Opportunity,* found parental education, income, and race to be strongly linked to student achievement with school factors being 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 into family factors 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)).

Many potential mechanisms may be at work in the SES-achievement connection (Cheng and Peterson (2019)). For example, 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 and adolescent experiences contribute to SES disparities in academic achievement (Kao and Tienda (1998); Perna (2006); Goyette (2008); Jacob and Linkow (2011)).

In empirical analyses, measures of SES are ordinarily based upon data availability rather than conceptual justification. In large-scale assessments of student achievement, data collection procedures usually ignore hard-to-measure qualitative family-related factors such as parent-child interactions, child upbringing approaches, or general physical and nutritional conditions (see, for example, Gould, Simhon, and Weinberg (2019)). Rather, the general approach is to look for more readily available indicators of persistent cultural and economic differences across families

as proxies for the educational input of families. The standard list includes parental education, occupation, earned income, and various items in the home (National Center for Education Statistics (2012a); Sirin (2005)). These measures tend to be highly correlated, making their separate impacts on learning and their relative importance difficult to disentangle.

While family income might generally be thought of as a good summary measure of SES, its reliability for this use has not been well validated and obtaining data on this from large-scale surveys is problematic. Survey data linked to assessments often come from the students themselves, and students generally have imperfect knowledge of their parents' earned income. For that reason, large-scale assessments that gather information directly from students seek to ascertain economic well-being by asking questions about consumption items, such as the number of durable and educational items present in the home. As compared to household earned income, students are intuitively better informed about whether a durable good (e.g., a dishwasher, computer, or a separate bedroom for themselves) is available in their home (Astone and McLanahan (1991), p. 313). Investigating the reliability and validity of student reports of parental SES characteristics compared to responses provided by mothers and fathers, Kayser and Summers (1973) conclude that "student reports were relatively stable over time and were more reliably measured for parental education than for either father's income or occupation. The validities of the reports were, for all but income reports, moderate. The validity of income reports was very low." An analysis by Fetters, Stowe, and Owings (1984) shows that student-reported indicators of parental education tend to be reliable but determined that "family income … was a matter of speculation for many students and thus inaccurately reported" (Kaufman and Rasinski (1991), p. 2). Consumption indicators may also be useful for estimating the resources of low-income families who supplement earnings with transfer payments, such as food stamps, medical services, housing assistance, and welfare benefits (Slesnick (1993)).

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

2.2 Trends in the SES-Achievement Relationship

Somewhat surprisingly, recent analyses have used income as their SES indicator. Reardon (2011b) has pioneered characterizing the SES-achievement gaps with income, and this is widely

cited in both academic literature and the general media.[3] His conclusion that income-achievement gaps have dramatically increased over the past half century is arguably the contemporary conventional wisdom.

Reardon (2011b) measures SES with data on household income obtained from students or parents. In analyses of data from twelve independent cross-sectional surveys, he estimates gaps in math and reading achievement of students at the 90th and the 10th percentiles of the household income distribution. 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, in his analysis, all of the change comes from changes in the 90-50 gap, whereas the 50-10 gap remains unchanged over the period—thus moving the discussion away from the historic concern about the troubles faced by disadvantaged students.

As we show in greater detail in Appendix B, there are two strands of evidence suggesting that these estimates largely reflect measurement error in both the achievement and income data used to estimate the SES-achievement gap over time. First, six of the twelve surveys that are used in the construction of SES-achievement gap data and that receive disproportionate weight in the trend analysis do not meet current quality standards for scientific surveys. They suffer to varying degrees from relying on student reports of parental income, having very large missing data problems, lacking appropriate achievement tests, and using very selective samples. When the trend analysis is limited to the more reliable surveys with reduced measurement error, no upward trend in the income-achievement gap is observed (Appendix Figure A1)—quite consistent with our findings here. Second, the two sets of data that employ psychometrically matched tests—NLSY79 and NLSY97 and ECLS-K, ECLS-B, and ECLS-K2010—do not show any upward trend in the income-achievement gap for a time period identified in the Reardon analysis as experiencing a steep increase. These findings also imply that the apparent difference in results between his study and ours is not due to the use of different SES measures—income vs. an index based on parental education and home possessions.[4]

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

[4] In combining test-gap information across surveys, Reardon (2011b) converts gaps on the various tests to standard deviations, assuming them to be equal; psychometricians have questioned treating as equivalent results from tests with different scales and effective domains (e.g., Lord (1950); Holland (2002); Dorans (2008); Ho (2009)). In addition, the gap estimates in Reardon (2011b) may be subject to measurement error from extrapolation from observed domains of the SES distribution to unobserved domains (see Appendix C).

Three other recent studies include trend data for the United States in international comparisons of SES-achievement gaps.[5] Although they differ in their data sets, methodologies, and operational definitions of SES, all three observe a diminution of the SES-achievement gap in the United States over the past two decades. The first study by the 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 that we also use. 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, described more thoroughly in Appendix A, resembles the SES measure used in our analysis (see section 4.2). Instead of looking at achievement gaps between the tails of the SES distribution, the study gauges changes in the SES-achievement connection by identifying 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 survey. The OECD (2018) reports a decline in R^2 over the fifteen-year period for the United States, which it interprets as indicating greater equity in the distribution of achievement.

The second study by 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 SES indicator, although she also separately analyzes parental occupation and books in the home. She finds no significant trend in the SES-achievement gap for the United States on the eight test administrations of student performance given to cohorts born between 1950 and 2001. There are, however, concerns about her treatment of tests that are not psychometrically linked and about her reliance on compressed parental-education categories to measure SES, requiring extensive extrapolation of achievement scores far outside the range of observed categorical SES information (discussed in Appendix C).

[5] Hedges and Nowell (1998) also report but do not analyze 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.

In the third analysis, Broer, Bai, and Fonseca (2019) employ psychometrically linked assessments in math and science administered by TIMSS to estimate trends in SES-achievement gaps for eleven countries including the United States between 1995 and 2015. 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).[6] They find that SES-achievement gaps in the United States decline significantly for science performance but do not change significantly for math.

Numerous 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 (2011b). These studies quite consistently identify a substantial closing of the black-white 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, changes in the black-white test-score gap may provide a partial window on trends in the SES-achievement gap. But the correlation between ethnicity 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.

3. Longitudinal Achievement Data

Building on these studies, we estimate trends in SES-achievement gaps 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. All are low-stakes tests: No consequences to any person or

[6] While more home resources are available, those employed in their study were restricted to these two in order to maintain comparability over time and across countries. Nonetheless, they compute the distribution of their SES index separately for each country-year observation. As their index is based upon a limited number of discrete SES category values that do not precisely match the 25th and 75th percentiles, they estimate the top and bottom quartiles by randomly sampling achievement values from adjacent categories.

entity are attached to student performances, and results are not identified by name for any school, district, teacher, or student.

All four surveys contain student background questionnaires that collect information about parents' education and about a variety of durable material and educational possessions in the home that we use in constructing an SES index. In addition, parental occupation is available in one survey, and student eligibility for free and reduced-price lunch is available from administrative records in two of the surveys. Each data set provides micro data at the student level that link questionnaire responses to students' test scores for each subject.

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

LTT-NAEP tracks performances of a nationally representative sample 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. LTT-NAEP data are available for reading in select years from 1971-2008 and for math from 1973-2008.[7] 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 2012. In a typical year, approximately 17,000 students participate in the administration of the LTT-NAEP. All NAEP data come from the National Center for Education Statistics (NCES) and were analyzed in a restricted-use data room.

Main National Assessment of Educational Progress (Main-NAEP)

Main-NAEP administers tests of math and reading aligned to the curriculum in grade 8.[8] 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

[7] LTT-NAEP also tests 9-year-olds, but we do not include these data in our analyses in part because of the limited, fragile information on SES background of the students. Also, our focus is 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 SES-achievement gaps for 1973. Thus, the achievement gap analysis that includes both math and reading is based upon two fewer observations than the level analysis.

[8] We exclude Main-NAEP science because 8th grade tests were administered in only two years, 2000 and 2005. As in prior research, we do not include results from exploratory surveys NAEP conducted prior to 1990 in part because the necessary information on SES is not publicly available. We also exclude other subject areas due to limited testing and uncertainties as to the accuracy of test measurement in these domains.

each participating state.[9] Main-NAEP maintains a reputation for reliability and validity similar to LTT-NAEP, and it was thought to track trends over time accurately enough that the LTT-NAEP was no longer necessary. For each administration of the test, the Main-NAEP sample is approximately 150,000 observations, the large sample being necessary in order to have representative samples for each state.

Trends in International Mathematics and Science Study (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 a limited number of countries.[10] The tests are designed to be curriculum-based and are developed by an IEA-directed international committee. Although early IEA tests were not psychometrically linked over time, beginning with the cohort born in 1981 (tested in 1995) the TIMSS tests have been designed to generate scores that are comparable from one administration of the survey to the next. We use the TIMSS 8th grade math and science tests beginning with this cohort. TIMSS data are available every four years from 1995-2015. The U.S. sample includes approximately 10,000 observations for each administration of the test.[11]

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 (rather than students at certain grade levels) 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,

[9] 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.

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

[11] We create a panel of the U.S. TIMSS micro data using national data files from 2003, 2007, and 2011, and international data files from 1995, 1999, and 2015. The only apparent difference between the 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.

so we use national PISA data available every three years from 2000-2015, though results are not available for reading for the 1991 birth cohort because of test administration problems.

4. Methodological Approach

We aggregate achievement and family background data from the four intertemporally linked surveys and construct an SES index similar to the one used by PISA in order to estimate trends in the SES-achievement gap over time.

4.1 Combining the Achievement Data Sets

We compile an aggregate distribution of achievement from student-level micro data 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 were administered to students between the ages of 13 and 15. The first test was administered by LTT-NAEP in reading to a cohort of students born in 1954; the last test was administered to students born in 2001. Across this near half-century span, achievement data are available for 2,737,583 students from 46 tests in math, 40 in reading, and 12 in science. Table 1 gives for each survey the number of assessments, subject matter, age or grade level at which students are tested, birth cohorts surveyed, and number of observations. Our overall sample contains 98 separate test-subject-age/grade-year observations.[12] Appendix Table A1 indicates the specific years in which the different surveys were administered.

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

To equate results across tests, we calculate achievement means and achievement gaps between groups in 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.[13]

[12] As indicated, on the first two LTT-NAEP math assessments, we only have mean achievement but cannot estimate SES-achievement gaps, reducing the number of analyzed test-administration observations to 96.

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

4.2 Measuring the SES-Achievement Gap

To estimate SES-achievement disparities, we need a consistent measure of students' SES. Here, we briefly describe the construction of our SES index, relegating details of the construction to Appendix A and discussing key methodological issues in section 6. Estimating achievement differences between different parts of the SES distribution requires a measure of SES that adequately depicts the full distribution of the population, rather than dividing it into a limited number of categories such as level of degree attainment of parents. Given the inaccuracy of student reports of parental income (and the ensuing lack of such data in large-scale assessments), we construct an index of SES based on student-reported information of parental education and home possessions, which is provided in all four assessment surveys included in this analysis.

We construct an SES index similar to the one used in PISA (OECD (2017a)). The index is given by the first principal component from a factor analysis of the two underlying variables—parental education and the number of home-possession items. Since the set of measured home possessions varies over time as does their individual utility for characterizing SES differences, we perform the principal component analysis separately for each test administration (for details, see Appendix A). These calculations allow us to capture the relative position of each child's family in the percentile distribution of SES in each given year.

Because of data limitations, we depart from the OECD's measure of SES by excluding 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 full 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 A2).[14]

Our measure of the SES-achievement gap is the difference in achievement between students in the top and bottom quartiles of the distribution of the SES index. That is, we compare the

[14] 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 (see Appendix A).

average score for the group of students at or above the 75th SES percentile to the average score for the group of students at or below the 25th percentile. For expositional purposes, we refer to this as the 75-25 SES gap. Appendix Table A2 reports this measure of the SES-achievement gap for each of the underlying test-subject-age-year observations, together with measures of the average achievement level and the overall achievement dispersion in the population.

The grouped SES-achievement comparison allows us to compare achievement across disparate broad segments of the population without imposing any functional relationship on the SES-achievement distribution. An alternative approach is to regression-estimate an empirical SES-achievement relationship and then to predict the achievement at specific SES percentiles from this estimated relationship, as done for example by Reardon (2011b), Chmielewski and Reardon (2016), and Chmielewski (2019) to estimate 90-10 gaps. As we show in section 6.4 below, we obtain similar results when using this point estimation approach to predict within the range of observed SES values. By contrast, extrapolation to extremes such as the 90th percentile is strongly sensitive to functional form assumptions (see Appendix C). Note, however, that our approach to look at the difference in the average performance between those in the top and bottom quartiles of the SES distribution means that the median student within each of these quartiles is located at the 87.5 and 12.5 percentiles, respectively—not far from the 90-10 extrapolations attempted in the other studies.[15]

In some of the underlying surveys, the available categorical information on the different SES components is somewhat limited (see section 6.3 below for details). To avoid measurement error arising from crude characterization of the SES distribution unsupported by the data, we exclude all test administrations where we cannot observe the portion of the SES group that falls in the relevant tail of the distribution corresponding to the desired SES comparisons. Specifically, 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 SES-achievement gap.[16] In practice, the concern is always about the top end of the SES distribution, because all of our observations have sufficient detail at the bottom end.

[15] If one were to assume a linear achievement function within the top and bottom quartiles, the median performance for the top and bottom quartiles would be the same as the group average, and the 75-25 gap reported here could be interpreted as broadly comparable to estimating 90-10 gaps by extrapolation.

[16] 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 use local linearization to interpolate achievement between the SES categories immediately above and immediately below the desired cut point.

This sample rule reduces our observations to 81 out of the potential 96 observations and postpones the first included birth cohort from 1954 to 1961, but it allows us to be more confident about identifying achievement patterns in the tails of the distribution. In other words, in making inferences about relative achievement in the tails of the distribution we do not want to impose fixed distributional assumptions on the achievement in the tails. 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.[17]

4.3 Estimating Trends in Achievement Gaps and Levels

While the four separate assessment regimes—LTT-NAEP, Main-NAEP, TIMSS, and PISA—are internally consistent over time, they vary from each other in a variety of details, including relationship to the curriculum, testing philosophy, and sampling frames. We assume that each testing regime provides 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 assessment regime, subject, and age group. The fixed effects for the four testing regimes take out any impact of regime-specific characteristics on the trend-line estimation.[18]

To estimate the trend 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:

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

where $\delta_i, \gamma_s,$ and λ_a are fixed effects for assessment regime, subject, and age; *t* is birth year; and ε is a random error. The parameters α_1 and α_2 describe the trend in achievement.

[17] For the same reason, we also refrain from estimating 90-10 gaps or comparable extremes of the distribution, as 45 of the potential 96 observations provide no way of distinguishing achievement of students in the top 10 percent of the distribution from those further into the distribution. In other words, in close to half of the observations the observed top category of the SES distribution has greater than 10 percent of the students.

[18] The fact that the aggregate trend line is estimated with a regression that makes use of information on trends within each separate psychometrically linked testing regime but not variations across the assessment regimes distinguishes the analysis from the Reardon (2011b) study, avoiding potential bias from scale differences.

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} :

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

In our main analysis, we focus on the SES-achievement gap as depicted by the achievement difference between those in the top and bottom quartiles of the SES index distribution, but we also report trends in other measures of disparities in our robustness analyses.

5. Main Results

We start by presenting results on the aggregate trend in the SES-achievement gap for all students in all subjects, followed by an exploration of heterogeneities by subject, age, and testing regime. We then report trends in the levels of achievement.

As a general background on achievement disparities in the U.S. student population, we note that the overall distribution of achievement, while narrowing a little, has shown only limited change. Figure 1 plots the trend in the achievement difference between students performing at the 75th and 25th percentiles of the achievement distribution, as well as between those at the 90th and 10th percentiles, over the past half century (birth cohorts 1954-2001).[19] 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 percentiles of the achievement distribution among those born in 1954 is close to 2.5 s.d.[20] Over the next half century, this gap (measured in units of the initial s.d.) closes slightly to about 2.3 s.d., indicating some shrinkage in the overall variance of achievement. The interquartile range in the achievement distribution is, almost by definition, smaller than the 90-10 dispersion. For students born in 1954, it is about 1.3 s.d. Over the next fifty years, the interquartile range declines modestly by 0.15 s.d.[21]

[19] Note that it is possible to get precise estimates of the bottom and top of the achievement distribution, because the overall performance distribution does not rely on characterizing students' SES, which is the cause of difficulty in estimating the tails of the SES-achievement distribution.

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

[21] This dispersion in the overall distribution of achievement varies slightly by subject matter (not shown). In math, both the 90-10 difference and the 75-25 difference close somewhat over the first half of the period but remain mostly flat in the second half. The dispersion in reading is 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. There is no difference by age group.

5.1 Trends in the SES-Achievement Gap

The trend in the 75-25 SES-achievement gap, as displayed in Figure 2, presents a startling picture: The connection between SES and achievement hardly wavers over the forty-year period (birth cohorts 1961-2001) for which we have sufficiently rich SES data.[22] The trend line, which is based on the within-test data and does not use the between-test data, is essentially flat. In the 1961 birth cohort, the predicted achievement gap between the average of those in the top and bottom quartiles of the SES distribution was 0.84 s.d. This gap increases only slightly over the course of the next forty years to 0.91 s.d. for the cohort born in 2001.

In the trend estimation, the linear and quadratic terms of the birth year are individually and jointly insignificantly different from zero in predicting the SES-achievement gap (individually, $p>0.90$ for each; jointly, $F(2,72)=0.53$ with $p=0.59$). This of course is not surprising given the flatness of the estimated trend line. The same is true if the birth year effect is estimated just linearly ($p=0.31$).[23] Inspection of confidence intervals, found in Appendix Figure A3, allows to visualize the size of the change in the SES-achievement gap that can be confidently rejected. While inference from the quadratic function necessarily becomes slightly imprecise at the endpoints of the observed time range, it is obvious that—despite the statistical uncertainty contained in the measurement of achievement on any one of the underlying assessments—we can confidently reject quantitatively noteworthy changes in the SES-achievement gap over the bulk of the observation period.

Trends are quite similar for math and reading performances (Figure 3).[24] The 75-25 gap for math narrows slightly over time. In reading, the disparity increases slightly over the entire period. The overall conclusion remains: the aggregate trend shown in Figure 2 does not mask different trends across subjects.

Trends in achievement gaps are also quite similar for younger and older students. Appendix Figures A4-A7 display all the observed underlying data points on the 75-25 SES-achievement gap (as well as on the 75-25 gap in the overall achievement dispersion) for each test regime,

[22] The quality of the SES measurement is noticeably poorer in the early survey years, implying that in going from the 96 potential observations to the more reliable 81 observations we lose the earliest cohorts of LTT-NAEP.

[23] Throughout, the quadratic form is employed to depict basic nonlinearities, but a linear form produces the same substantive conclusions on the trends in SES-achievement gaps.

[24] For clarity, we leave out the individual data points in all depictions except Figure 2. Markers on the trend lines indicate birth cohorts for which there is data.

subject, and age.[25] The trends in the SES-achievement gap at ages 14 and 17 in the LTT-NAEP (Appendix Figure A4) do not show different patterns at the two age levels.

Inspection of the scatter of data points on Figure 2 and the respective appendix figures shows that with the exception of PISA each of the trends within testing regimes resembles the overall trend. As previously indicated in the OECD (2018) analysis, PISA gaps show a sizeable closing of SES gaps for cohorts born between 1985 and 2001 of about 0.35 s.d. in math and 0.4 s.d. in reading (although not in science). If PISA observations are excluded from the aggregate trend (leaving 64 observations), the gap increases from 0.86 to 0.99 for cohorts born between 1961 and 2001. If we exclude each of the other three test regimes one at a time and re-estimate the trend, the lines are essentially flat, with the joint quadratic parameters being insignificantly different from zero for each subsample. We have no reason to question the validity of any of the separate testing regimes and therefore consistently rely on trends aggregated from the within-regime time patterns by birth cohort.

There is the possibility that trends differ in other parts of the distribution. In Reardon (2011b)'s analysis, for example, the increasing trend is observed only in the upper half of the distribution. To see whether our data yield a similar trend, Figure 4 compares the top quartile with those in the bottom half of the SES distribution (75-50 gap) and the top half with the bottom quartile (50-25 gaps). While the trend line for the 75-50 gap curves slightly upwards and the 50-25 gaps slightly downward, the estimated quadratic coefficients for both gaps are insignificantly different from zero. A similar overall pattern emerges when—instead of taking the top and bottom halves of the SES distribution—alternatively we calculate achievement for students in the 45-55 percentile band of the SES distribution and consider the 75-50 and 50-25 gaps using this estimate for the mid-range of achievement (Appendix Figure A8).

The finding of little change in SES-achievement gaps can be viewed from two different vantage points. On the one side, we do not perceive any trend similar to that reported in Reardon (2011b), which shows large increases in gaps between the two extremes of the distribution. On the other side, we do not find any overall narrowing in achievement gaps despite a half-century of educational programs designed to win a "war on poverty."

[25] There is some instability in the SES gap estimates for the individual assessments by subject. The individual figures connect available data points even if various intermediate points are missing because of inability to isolate performances among those above the 75th percentile of the SES distribution.

5.2 Trends in 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 longitudinal data on student outcomes, we can directly evaluate whether there are gains in student achievement and, if so, whether they are general or isolated gains.

Figure 5 reports the nonlinear trend in achievement levels, as estimated by equation (1). For adolescents who are approximately age 14, we observe a sharp increase of about 0.46 s.d. over the time period or approximately 0.09 s.d. per decade.[26] By contrast, total 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]

There is significant heterogeneity in the trends in achievement levels by subject. Mean achievement gains by cohorts are largely concentrated in math. Younger adolescents register a math improvement of about 0.67 s.d., while the older ones show an overall shift upward of about 0.2 s.d. (Appendix Figure A9, Panel A). Reading gains are smaller: The trend among older adolescents shows no improvement, while the trend among younger adolescents amounts to only about 0.23 s.d. over the half century (Panel B).

Importantly, the trends in SES-achievement gaps considered earlier are essentially the same for both the younger and older students. As shown in section 5.1, we detect very little temporal change in achievement gaps in both cases regardless of subject matter.

---

[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. To see this, we estimate the LTT-NAEP scores for 2012 as if the population had the same ethnic distribution as in 1980. We re-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.

6. Methodological Issues and Robustness

The clear pattern of SES-achievement gaps conflicts both with past work and with common perspectives on compensatory policy actions. To show its robustness, we trace through a variety of methodological and data issues. We begin by showing that our main findings are confirmed when estimated based on student eligibility for subsidized lunch programs and when considering changes in the ethnic composition of the population. We then address several methodological issues in estimating SES-achievement gaps from categorical data, showing robustness to alternative methodological choices. We discuss limited information in some observations of the SES tails and show robustness to an expansion of test observations when considering the 70-30 rather than the 75-25 SES gap. We also show robustness in an alternative point estimation approach to estimating SES gaps. Relatedly, we discuss the measurement error from extrapolating beyond the observed SES range in Appendix C. Finally, we show that basic conclusions are unchanged when treating achievement data as ordinal, i.e., only having rank-order interpretation.

6.1 Achievement Gaps by Eligibility for Free and Reduced-price Lunch Programs

Participation in federal lunch programs provides an alternative, income-based proxy to measure students' SES. Data available in the Main-NAEP allow us to estimate gaps in performance between students who are eligible and those who are not eligible for participation in the federal free and reduced-price school lunch programs at school for cohorts born between 1982 and 2001. Students who come from households at or below 130 percent of the poverty line are eligible for free lunch (whom we refer to as extremely poor), while those from households between 130 and 185 percent of the poverty line are eligible for participation in the reduced-price lunch program (whom we refer to as poor). In Main-NAEP, the indicator of eligibility for these federal programs is obtained from administrative records.

Analyzing eligibility for subsidized lunch 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 participating 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 years 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 percent (Chingos (2016); Greenberg (2018)). For these reasons, we regard this variable as only a crude SES indicator, but it does provide a direct (albeit imperfect) measure of household income over part of the time period.

When measured by subsidized lunch eligibility, the gap between the extremely poor students and other students in the 1982 birth cohort is a sizeable 0.71 s.d. (Figure 6). When the extremely poor are combined with the poor 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 one of these federal programs narrows by 0.01 s.d. Just like the results based on the SES index, this measure of the 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 above. Nonetheless, these results are entirely consistent with the trends for the 75-25 SES-achievement gap.

6.2 Black-White Achievement Gaps and Consideration of Ethnic Composition

As discussed at the end of section 2.2 above, changes in the ethnic composition of the U.S. student population mean that changes in the black-white gap can provide only a limited picture of changes in the SES-achievement gap. But separate estimation of changes in the SES-achievement gap for white and black students allows us to show robustness of our main finding to consideration of changes in the ethnic composition.

Black-white achievement gaps. We use NAEP data to estimate trends in the black-white test-score gap. 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.[29] 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.

[29] PISA does not make race information publicly available, and TIMSS, which collects information on race from student questionnaires, does so for only a subset of its survey administrations.

Our results on the black-white test-score gap in Figure 6 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 (2011b) that shows a strong decline 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 black-white 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 black-white achievement gap and the SES-achievement gap have remained essentially flat.

Changes in ethnic composition. Some have hypothesized that the lack of success in diminishing the size of the SES gap is due to changes in the 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 aged 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 (U.S. Department of Education (2013), Table 20).[30]

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 7). These calculations use the overall national SES distributions for both (i.e., the same as used in the aggregate 75-25 SES analysis above). Interestingly, the SES gaps for both whites and blacks increase modestly by about 0.1 s.d. over the past twenty years. This increment occurs at a time when the percentage of both whites and blacks in the upper ranges of the SES distribution has increased (Appendix Table A3). The percentage of blacks in the bottom 25 percent of the distribution has declined even more rapidly than the white percentage. Altogether, there is little evidence that changes in the ethnic composition of student cohorts account for the unwavering SES-achievement gap.

[30] 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.

6.3 Limits of Categorical SES Information and Extended Observations for 70-30 Gap

When estimating gaps between groups at the tails of the SES distribution, one needs response categories to survey items that distinguish between those who are at the tail of the distribution from those who are not. Otherwise, serious potential for measurement error may be introduced by including observations that are outside the portion of the distribution under investigation. As indicated in section 4.2 above, the SES index that we use to estimate the SES-achievement gaps is constructed from survey information on parental education and home possessions (see Appendix A for details).

There are three underlying aspects of the survey data that lead to 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, because of limits on the depth and breadth of any set of questions, typical survey designs provide the most detailed information about those near the middle of the distribution of the population rather than those at the extremes. As a consequence, in particular those at the top end of the distribution are not well distinguished from those located more toward the middle. Third, individual categories of the categorical questions may include a large percentage of all observations. The limited categorization adds to the difficulty of distinguishing different parts of the SES distribution on any derived SES index, again becoming particularly severe at the top end of the SES distribution.

In the data sets underlying our analysis, 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. A couple of examples chosen for illustrative purposes can document the range of available detail in SES measurement.[31] In 1990, the first year for Main-NAEP, students were asked to place their parents within one of four education categories and whether they had each of only four items in

[31] Appendix A provides a fuller description of the range of parental-education and home-possession information available in the separate test administrations. In general, there is little variation in the measurement of parental education but more variation in the measurement of home possessions. However, most of the variation is between assessment regimes—with richer home-possession information in TIMSS and PISA than in LTT-NAEP and Main-NAEP—rather than within assessment regimes. These between-regime differences are taken out in our trend estimation that contains regime fixed effects. To ascertain that differences in the number of parental-education categories and home-possession items do not affect our overall results, we included both counts as control variables in our trend estimation of the SES-achievement gap. Neither the number of parental-education categories nor the number of home-possession items enters the trend estimation significantly, and the estimated birth-year trend remains statistically insignificant (not shown).

their home (see left panel of Appendix Table A4 for specific survey items). As a consequence, when we construct our SES index based on these two measures, no less than 26.5 percent of the students that Main-NAEP tested in 1990 are identified as being in the top SES category, making it difficult to obtain a precise estimate of those in the top quartile of the distribution. Similar problems emerge for other surveys in early administrations of both the Main-NAEP and LTT-NAEP.

This limited categorical information contrasts sharply with other survey implementations. For example, the PISA survey in 2015 inquired about seven parental-education categories and the presence of twenty-two items in the home (right panel of Appendix Table A4). As a result, just one percent of the sample falls into the top SES category in the PISA 2015 survey. Figure 8 shows the frequency distribution for distinct categories identified along the continuum of our constructed SES index for the two example surveys. The tall spike at the top end of the Main-NAEP 1990 SES distribution represents the 26.5 percent of the sample falling in the top SES category that can be observed. The more granular description of the SES distribution in the PISA 2015 data derives from the greater number of parental-education and home-possession categories in that survey.

These examples illustrate the reason why our main analysis excludes any test administration where more than a quarter of observations fall in the top category of our SES index, reducing the number of included SES gap observations from 96 to 81 (section 4.2). That is, the Main-NAEP 1990 observation just described is *not* included in our analysis above.

To investigate whether the gap trend estimation is sensitive to the reduction in underlying observations, we estimate gaps for the slightly expanded tails of the 70^{th} and 30^{th} percentiles of the distribution, which are identified in several additional test administrations (including the Main-NAEP 1990 example) compared to the 75^{th} and 25^{th} percentiles. In particular, analysis of the 70-30 distribution allows us to reliably observe the relevant tails of the distribution in 91 (of the potential 96) observations.

Analysis of the expanded set of observations on the 70-30 gap confirm our main results for the 75-25 gap. The markers on the trend lines in Figure 9 indicate the years for which an assessment is available, showing the increase in information from 81 to 91 observations. The 70-30 gaps are necessarily slightly smaller than for the interquartile range. More importantly, the trend line for this expanded sample remains essentially flat. Both individually and jointly, it

is again not possible to reject the null that the coefficients for the linear and quadratic terms in the trend equation are zero ($F(2, 82)=1.05$).

6.4 Group Calculations vs. Point Estimation of Gaps

Our main analysis provides a grouped SES-achievement comparison based on estimates of the average performance of students within a specific segment of the SES distribution. In particular, 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 (section 4.2). By focusing on this comparison, we do not have to characterize the precise pattern of achievement within the extremes of the SES distribution.

An alternative approach—used, for example, in Reardon (2011b), Chmielewski and Reardon (2016), and Chmielewski (2019)—is to compare achievement at specific points in the SES distribution, such as the estimated achievement of somebody exactly at the 25th percentile or the 75th percentile. This approach involves two steps. The available data provide the average achievement for a given category of SES values, i.e., for a range of SES percentiles. The first step is to identify a specific SES percentile that corresponds to the average score observed in each SES category. While there is no information within the assessment data that would guide this choice, a convenient approximation used in the prior analyses is to assume that the midpoint of the SES percentile range corresponds to the average achievement of that SES group (which amounts to a linearity assumption within each SES category). For example, for the 26.5 percent in the top SES category of the 1990 Main-NAEP distribution (discussed in the previous section), the average achievement is assumed to be an accurate estimate of the performance of students at the 86.75th percentile. The second step employs a linear or cubic regression function to estimate the SES-achievement relationship through all available SES data points observed for each test-subject-age-year observation. This approach assumes that all of the data points in the center of the SES distribution are useful in predicting achievement in the tails of the distribution. With this regression function, achievement is predicted at the respective SES percentiles of the gap under consideration.

As a robustness check, we implement this alternative point estimation approach to estimate 75-25 SES-achievement gaps in our data. We use an estimated cubic function for each test-subject-age-year data to define the relevant SES-achievement gaps. We predict achievement at the 75th and at the 25th percentiles of the SES distribution and calculate the gap between the two.

We also use this method to expand the analysis to all 96 available test-subject-age-year observations. As Figure 10 shows, the point estimation approach yields similar results to our main approach. The estimated 75-25 SES gap is essentially flat over time, with a very small downward trend that is statistically and quantitatively insignificant. As these gaps refer to the achievement difference between individuals at the 75th and 25th percentile of the SES distribution, rather than to the difference between the average achievement of those above the 75th and below the 25th percentile, their magnitude is obviously lower compared to the main analysis.[32]

6.5 Ordinal Analysis of Achievement Data

Prior research on achievement gaps most frequently treats test-score information as interval data. With that assumption, numerical differences in test scores at any point in the distribution, usually expressed in standard deviations, can be treated as equivalent to one another. However, some research has suggested that this assumption can lead to over-interpretation of achievement differences from standardized tests and has advised relying on an ordinal (rank-order) interpretation of test scales instead (e.g., Ho (2009); Bond and Lang (2013); Nielsen (2015)).

To understand better the potential impact of an ordinal approach versus the cardinal approach we previously employed, we consider illustrative examples of trend analysis in SES-achievement gaps over time using only ordinal, rank-preserving assumptions of test-scale information.

We again distinguish two groups of students, those in the bottom and top quartiles of the SES distribution, and now create the score distributions for both groups. For each percentile of the achievement distribution of the low-SES group, we can calculate the share of students in the high-SES group whose achievement is at or below the low-SES percentile's indicated achievement.[33] We plot these two achievement distributions against each other. If there was an equal achievement distribution for the top and the bottom quartile of the SES distribution, this

[32] For the reasons discussed in the previous section, we refrain from using the point estimation approach to estimate 90-10 SES-achievement gaps. Such analysis requires extrapolating points far outside the range of observed data such as extrapolating the precise value of expected achievement at the 90th percentile of Main-NAEP 1990 when we just observe the average achievement for students in the range of 73.5-100. Appendix C describes the extrapolation problem in greater detail.

[33] This approach corresponds to analysis of probability-probability plots described in Ho (2009).

plot would appear as a 45-degree line—just as with a Lorenz curve.[34] The greater the divergence from this line, the greater the inequality. Performing the same analysis on tests at different points in time allows for an assessment of the change in inequality over time that does not depend on interval interpretations but uses just the rank information in the tests. As the analysis requires a clear distinction of the two SES groups at the respective SES quartiles, we perform the analysis for the PISA and TIMSS tests, but—for the reasons indicated above—refrain from analysis of the NAEP data with this approach because of the lumpiness of the SES distributions.

These ordinal analyses yield conclusions very similar to our main analysis reported above. For example, using the earliest and latest installment of the PISA test, the top panel of Figure 11 shows that the 2015 distribution is considerably more equitable than the 2000 distribution, as the distance of the curved line from the 45-degree line for that year encloses much less space than the curved line for 2000. The implied reduction in inequality confirms the results of our main analysis for PISA math gaps discussed in Section 5.1 and shown in Appendix Figure A7. The bottom panel of Figure 11 shows a similar change in the inequality when comparing students in the top half to those in the bottom quarter of the SES distribution. Using the TIMSS data, Figure 12 indicates a slight increase in inequality between 1995 and 2015, just as had been implied by the interval analysis shown in Appendix Figure A6. Thus, for both PISA and TIMSS, the ordinal analysis that treats the assessment data as ordinal rankings confirms the trends in inequality estimated in our main analysis that assumes interval interpretability of the underlying scores.

7. Conclusions

Our analysis of long-run trends has shown that performance disparities within the United States are both large and highly persistent. The SES-achievement gap remains essentially as large as in the mid-1960s 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 policies were the centerpiece. In terms of learning, students in the top quarter of the SES distribution continue to be around three years ahead of those in the bottom quarter by eighth grade.

[34] Note that other properties of this curve differ from Lorenz curves. For example, it is entirely possible to have points above the 45-degree line if the share of high-SES students who score below a certain achievement threshold exceeds the equivalent share of low-SES students.

The constant disparity is not relieved by rising levels of achievement for all students at the end of their secondary education. Students in their early adolescent years show achievement gains over the past half century, making them better prepared for entry into high school. But at least over the past quarter century, these achievement gains disappear by the age of 17, just as students reach the point of entering college or the labor market.

Our findings are consistent with a variety of combinations of demographic changes and policy shifts that have occurred over the past half century. The flat trend could be due to the absence of relative change in the family and school inputs for students in the top and bottom parts of the SES distributions, due to offsetting trends in society that counter-balance one another, due to growing inequalities in society offset by equalizing policies, or due to equalizing trends in society offset by inegalitarian policies. While the aggregate trend data currently available do not allow us to provide definitive evidence to choose among these possibilities, let alone identify the underlying causal relationships, we can summarize some of the potential mechanisms suggested in the literature.

On the family side, differential trends in inputs at the top and bottom tail of the SES distribution are conceivable. On the one hand, there is a trend towards increasing disparity in household income and wealth within the United States, in particular at the very top end of the distribution (e.g., Krueger (2003); Autor (2014); Saez and Zucman (2016); Alvaredo et al. (2017)). In addition, SES differences in the incidence of single-parent households and in the age of mothers at the birth of the child may have increased over time (Duncan, Kalil, and Ziol-Guest (2017)). On the other hand, SES differences in parental education and in the number of siblings in the household—two factors consistently identified as important determinants of student achievement—have narrowed over time. Improvements in nutrition, health, and general economic well-being may also have disproportionately occurred in low SES households.

Differential trends are also conceivable on the policy side. Obviously, a long list of programs has been introduced with the intention of closing SES-achievement gaps. These programs include, for example, racial school desegregation following the 1954 Supreme Court decision in *Brown v. Board of Education*, particularly in the South (e.g., Rivkin and Welch (2006), Rivkin (2016)); compensatory funding of schools under Title I of the Education and Secondary Education Act of 1965 (e.g., Cross (2014)); expanded special education programs starting with the 1974 Education for All Handicapped Children Act (e.g., Morgan, Farkas,

Hillemeier, and Maczuga (2017)); state court decisions mandating greater fiscal equity (e.g., Peterson and West (2007), Hanushek and Lindseth (2009), Jackson, Johnson, and Persico (2016), Lafortune, Rothstein, and Schanzenbach (2018)); significant early childhood programs with the federal Head Start program and parallel state-funded programs (e.g., Friedman-Krauss et al. (2018)); and introduction of test-based accountability emphasizing disadvantaged students in the No Child Left Behind Act (e.g., Hanushek and Raymond (2005); Peterson (2010); Figlio and Loeb (2011)). What is less clear from available evaluations is how successful these programs have been in reducing the SES-achievement gap. Also, other trends in school inputs may have worked to increase achievement gaps between the top and the bottom of the SES distribution, such as segregation of schools by SES lines, school support programs from wealthy parents, or personnel policies that discourage the presence of high-quality teachers in schools serving low-SES students (Duncan and Murnane (2014)).

We cannot resolve the relative importance of the countervailing trends in family and policy inputs here. Our goal is just to clarify the pattern of SES-achievement gaps that guides many policy discussions. On the one side, we reject the often-raised claim that the SES-achievement gap is widening in the United States. We find almost no evidence at all for this proposition. On the other side, suggestions that the SES-achievement gap is closing are premature. While PISA data, considered in isolation, may be cited to support such claims, its findings are not corroborated by results from the other three test regimes, each with high-quality psychometrically linked assessments of student performance. We also cannot explain the puzzling pattern of differing trends of achievement levels at different ages, in particular because previous research rejects a number of methodological explanations (Blagg and Chingos (2016)).

The bottom line of our analysis is simply that—despite all the policy efforts—the gap in achievement between children from high- and low-SES backgrounds has not changed. If the goal is to reduce the dependence of students' achievement on the socio-economic status of their families, re-evaluating the design and focus of existing policy programs seems appropriate. As long as cognitive skills remain critical for the income and economic well-being of U.S. citizens, the unwavering achievement gaps across the SES spectrum do not bode well for future improvements in intergeneration mobility.

References

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

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

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

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

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." *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.

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

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

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

Cheng, Albert, and Paul E. Peterson. 2019. "Experimental Estimates of Impacts of Cost-Earnings Information on Adult Aspirations for Children's Postsecondary Education." *Journal of Higher Education* 90, no. 3: 486-511.

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

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

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

Duncan, Greg J., and Richard J. Murnane. 2014. *Restoring Opportunity: The Crisis of Inequality and the Challenge for American Education*. Cambridge, MA: Harvard Education Press.

Duncan, Greg J., and Richard J. Murnane, eds. 2011. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage Foundation.

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

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

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

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

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

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

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

from Death, Divorce, and Family Size." NBER Working Paper 25495. Cambridge, MA: National Bureau of Economic Research (January).

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

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

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

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

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

Hanushek, Eric A., 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, Laura M. Talpey, and Ludger Woessmann. 2019. "The Unwavering SES Achievement Gap: Trends in U.S. Student Performance." NBER Working Paper 25648. Cambridge, MA: National Bureau of Economic Research (March).

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: 15-19.

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

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

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

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.

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

Holland, Paul W. 2002. "Two Measures of Change in the Gaps Between the CDFs of Test-Score Distributions." *Journal of Educational and Behavioral Statistics* 27, no. 1: 3-17.

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

Jacob, Brian A., and Tamara Wilder Linkow. 2011. "Educational Expectations and Attainment." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation: 133-162.

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

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

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

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

Krueger, Alan B. 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.

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

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

OECD. 2017a. *PISA 2015 Results (Volume III): Students' Well-being*. Paris: Organisation for Economic Co-operation and Development.

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

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

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

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

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

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

Reardon, Sean F. 2011a. "Appendix to: The widening academic achievement gap between the rich and the poor: New evidence and possible explanations." (mimeo) Stanford, CA: Stanford University.

Reardon, Sean F. 2011b. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In *Whither Opportunity? Rising*

Inequality, Schools, and Children's Life Chances, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation: 91-116.

Reardon, Sean F., and Ximena A. Portilla. 2016. "Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry." *AERA Open* 2, no. 3: 1-18.

Rees, Daniel I., and Joseph J. Sabia. 2010. "Sports participation and academic performance: Evidence from the National Longitudinal Study of Adolescent Health." *Economics of Education Review* 29, no. 5: 751-759.

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.

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

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

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

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

U.S. Department of Health and Human Services. 2018. NICHD Study of Early Child Care and Youth Development: Phase I, 1991-1994: Inter-university Consortium for Political and Social Research.

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

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.

Appendix A: 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 interquartile range of the SES distribution, 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-price 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.

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

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

To compute the ESCS index, PISA uses a combination of highest parental education (in years), parental occupation (transformed into an International Socio-Economic Index of Occupational Status (ISEI), see Ganzeboom and Treiman (2003)), and home possessions (derived from ten to fifteen yes/no questions such as "Do you have a desk to study at in your home?" and three to five questions such as "How many cars do you have at your home? (None, one, two, three or more)", see Appendix Table A4 for further examples). PISA standardizes the three variables, performs a 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 A11, 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 (Panel B).

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 make adjustments to synchronize with the other surveys in our analysis that do not include all the information available in 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.

In all LTT-NAEP and Main-NAEP waves, parental education is measured in four categories (see Appendix Table A4). PISA always uses seven categories. TIMSS mostly uses five

categories, with the exceptions of the first two waves, where it is three and four categories, respectively.

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

Home possessions. To create ESCS, the OECD uses an index of home possessions (*homepos*) which is "a summary index of all household and possessions items" (OECD (2017b)). NAEP and TIMSS include similar questions about students' home possessions, but they do not provide a summary index. 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 report owning (across both dichotomous and categorical questions) and use this number as our *homepos* variable in the specific survey and year.[35]

There is some variation in the number of available categories of home-possession items, in particular between the two NAEP assessments and the two international assessments but also within assessment regimes over time. In LTT-NAEP, the number of home-possession items is four in the first two 1970s waves, up to eleven in the next two waves, down to four and six in the subsequent two, and steady at seven ever since 1998. In Main-NAEP, it is between four and six in the 1990s, between five and eight in the 2000s, and between six and nine in the 2010s. The available information on home-possession items is much larger in TIMSS and PISA throughout. In TIMSS, it is twenty in the two waves of the 1990s and between eleven and thirteen in the subsequent four waves. In PISA, it is 38 in 2000, 28 in 2003, and between 44 and 46 in the

[35] 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.

subsequent four waves (all reported numbers refer to the sum of possible item counts across the dichotomous and categorical questions; see Appendix Table A4 for examples).[36]

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

In the construction, we differ slightly from the ESCS process in the treatment of missing variables. The OECD treats missing variables in the following way: "For students with missing data on one out of the three components, the missing variable was imputed . Regression on the other two variables was used to predict the third (missing) variable, and a random component was added to the predicted value. If there were missing data on more than one component, ESCS was not computed and a missing value was assigned for ESCS" (OECD (2017b)). As this method requires the assumption of a positive, linear relationship among the variables and in any case only applied to 2 percent 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 tails of the distribution, we compare trends in the top and bottom quartiles, respectively, 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 A2).

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

[36] Note that because our main analysis excludes test administrations where the top category of the constructed SES index includes more than a quarter of the population, most of the test administrations with very few home-possession items are not included in our main analysis. In particular, with the exception of only one test administration that has four home-possession items, all test administrations included in our main analysis have at least six home-possession items.

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 parental education and home possessions).

The correlation between the SES index and reported family income is displayed in Appendix Table A6. 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.[37]

[37] 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.

Appendix B: Comparison with Reardon (2011b)

Our results differ from those reported by Reardon (2011b), who finds rapidly growing SES gaps when SES is measured with parental income indicators. In this Appendix, we reconcile our findings with his by showing that the difference in results is due to systematic measurement errors in his analysis. When a trend line is estimated from the studies in his analysis less prone to measurement error, no upward trend in the SES-achievement gap is detected.

Reardon estimates trends in the income-achievement gap for cohorts born between 1942 and 2001 from observations of average math and reading test performance obtained from "twelve nationally representative" samples of students of various ages. Appendix Table A5 provides a list of the included databases and their acronyms which we use here. The underlying databases were selected by Reardon because they were accompanied by surveys that included student- or parent-provided data on household income (p. 93) along with student achievement test information.

Within each data set, Reardon computes a value of the achievement gap between students at the 90th percentile of the income distribution and those at the 10th percentile from a cubic regression of achievement on parental income.[38] He treats every age cohort within each test administration as an independent observation, giving him 20 estimates of math performance gaps and 26 of reading performance gaps. These are the data points for the subsequent trend analysis.

He uses a quadratic equation to estimate the trend for these SES-achievement gaps by subject over the entire time period and a linear equation to estimate the trend between 1974 and 2001. He finds that the "achievement gaps [between those at the 90th and the 10th percentiles of the income distribution] among children born in 2001 are roughly 75 percent larger than the estimated gaps among children born in the early 1940s" (p. 95).

Reardon (2011b) himself expresses concern "that the trend in the estimated gaps for the earliest cohorts, those born before 1970, is not as accurately estimated as the later trend. … Family income was reported by students rather than by a parent … [and they] are school-based samples of students in high school [that] exclude dropouts, who are disproportionately low-income and low-achieving" (pp. 95-96). He is more confident that the trend has shifted upward for cohorts born between 1970 and 2001. Between 1974 and 2001, "the income achievement

[38] Issues related to extrapolating the observed SES percentiles are discussed in Appendix C and present an additional component of measurement error.

gap has grown by roughly 40 to 50 percent …, a very sizeable increase" (p. 97). Elsewhere, he describes this increase as "roughly 30 to 40 percent" (p. 93). Unfortunately, his analysis does not reflect the overall impact of the data problems, many of which he himself identifies in an appendix (Reardon (2011a)).

There are two strands of evidence indicating that the perceived trends are not real but instead are a function of measurement error. First, in a significant portion of constructed gaps household income is very poorly measured; in another portion the test instruments themselves are quite suspect; and in a final set survey sampling and missing data become serious issues. When surveys at high risk of serious measurement error are excluded, no upward trend in the income-achievement gap is observed. Second, comparisons of gap differences among psychometrically matched observations that cover the relevant period of supposed steep gap increases reveal no upward trend.

Data Quality Issues

Survey researchers have not found it easy to collect accurate data on family income, the independent variable in the regressions that Reardon uses to create the gap estimates for each survey. As discussed in section 2.1 above, the challenge is especially large when household income is estimated from student-provided information (Kayser and Summers (1973); Fetters, Stowe, and Owings (1984); Kaufman and Rasinski (1991)).

In Reardon's analysis, seven of the eight earliest estimations of the income-achievement gap are based upon "family income … reported by students rather than by a parent" (p. 95). Given the dependence on student reports, it is almost certain that these seven observations measure the income-achievement relationship with much greater error than the subsequent studies that collect income information directly from parents. In the construction of observations for SES-achievement gaps, these errors will bias the income-achievement relationship toward zero and will yield downward-biased data points on gaps. The improvement in data collection techniques over time then contributes to the appearance of a rising income-achievement gap when none exists.

Six of his twelve surveys are plagued with serious problems and clearly do not meet current scientific quality standards. Importantly, these surveys introduce systematic measurement error into the subsequent trend estimation based upon them. From the information provided in

Reardon (2011a, 2011b) and the online descriptions, these surveys do not provide reliable data points but nonetheless have a decided influence on the shape of the estimated trend line.

Project Talent

Talent, the earliest survey, provides estimated gaps for four cohorts born in 1942-1945. It uses student-provided estimates of family income, employing five income categories.[39] While there are questions about the sampling in the Talent survey, the largest concerns relate to nonresponse rates to the question concerning family income. No less than 54 percent of the freshmen, 50 percent of the sophomores, 45 percent of the juniors, and 39 percent of the seniors chose not to "guess"—the word used in the survey—the answer to the family income question.

For the early Talent cohorts born in the early forties, Reardon reports income-achievement gaps in reading and math of approximately 0.75 s.d. Only Prospect, which has its own measurement problem (see below), reports gaps of a similarly low magnitude.

National Longitudinal Study of 1972 (NLS)

The National Longitudinal Study of 1972 was an early survey conducted by the National Center for Education Statistics. It was designed to follow a sample of high-school seniors of the Class of 1972 into the labor market and college. Like Talent, its parental-income measure came from sampling the students, who place estimated income into one of ten categories.[40] Twenty-one percent of the respondents chose not to respond to this question.

High School and Beyond (HS&B)

The High School and Beyond study collected income data for birth cohorts in 1962 and 1964. An important element of these surveys was a 15 percent sample that collected income data from both a parent and the student. These data were particularly important in the analysis because they provided estimates of the impact of student-provided income data. Unfortunately, the parent-provided income data lacked face validity, being considerably above the comparable

[39] As discussed in Appendix C, the impact of having just five categories depends on where these categories cut the income distribution. Reardon (2011a) considers the impact of differing numbers of categories that are evenly spaced across the income distribution. Actual survey data on categorical income, however, is almost certainly going to have unevenly spaced data categories with limited observations in the tails of the distribution.

[40] The top household income category in the NLS survey was income greater than $18,000. From the CPS, 16 percent of households had incomes that would be included in this top category.

values from the Current Population Survey (CPS).[41] Reardon uses these parent-provided data to estimate reliabilities for the student-provided data. He concludes that the student data are more reliable than the parent data. Partly because of this and partly because of sample sizes, he uses the student-provided data from HS&B to estimate the income-achievement relationship and the subsequent SES-gap points for the trend analysis.

To get some sense of the impact that measurement error can have on estimates of the income-achievement gap, one can compare HS&B estimates to those observed in NLSY79, which was administered to birth cohorts of students that overlapped with those observed in HS&B.[42] NLSY79 is much less likely than HS&B to err in its estimation of household income. The highly-regarded NLSY79 protocol, administered to parents, obtains household income as a continuous variable with just a 2 percent non-response rate. In contrast, the student-provided HS&B data are categorical and have non-response rates of 13 and 18 percent of seniors and sophomores, respectively. Because of the differential measurement error, it is not surprising that the average of the math and reading income-achievement gaps identified in the NLSY79 survey is 1.3 s.d., which is fully 0.35 s.d. higher than the gap reported in the HS&B survey.

The Congressionally Mandated Study of Educational Growth and Opportunity (Prospects)

Problems with the seldom-used Prospects survey are noted in Reardon (2011b), p. 112, note 70: "It is difficult to find documentation on the content and psychometric properties of the Prospects tests. These tests may be much less reliable than other tests; as a result, I am inclined to discount their importance in describing the trends." Contrary to that statement, Reardon uses three observations from Prospects to estimate trend lines in both math and reading, while six other studies yield only one observation.[43]

The estimates of the income-achievement gap from Prospects vary widely across the three cohorts surveyed at the same time—by 0.3 s.d. in math and 0.35 s.d. in reading. All three of the

[41] Reardon (2011a) reports the difficulties with the complex survey: "In HS&B, parent-reported family income is measured using a set of survey questions, rather than a single question, as in other studies. The responding parent—usually the mother—was asked 1) how much wage income s/he received; 2) how much self-employment income s/he received; 3) how much wage income his/her spouse received; 4) how much self-employment income his/her spouse received; and then 5) a set of 15 questions asking how much the respondent and spouse together received from other sources, including dividends, interest, rent, alimony, AFDC, SSI, etc." As a result, average income from the sum of parental responses is 33 percent above that reported in the CPS at the 90th percentile.

[42] NLSY79 surveyed cohorts born in 1961 and 1963, while HS&B surveyed those born in 1962 and 1964.

[43] Only one birth cohort is observed with data from NLS, NELS, ELS, SECCYD, ECLS-K, and ECLS-B.

Prospects estimates are much lower than the 1.3 s.d. gap obtained from NLSY97 administered at about the same time. In other words, estimates for the income-achievement gap for cohorts born within a couple of years of 1980 vary by no less than 0.6 s.d. Measurement error is the most likely explanation for this spread in estimates made at close to the same time.

National Longitudinal Study of Adolescent to Adult Health (Add Health)

Add Health contains no math achievement data, but the study contributes over a third of the post-1970 data points used by Reardon to estimate the trend in the reading gap. The observations cover students at the ages of 13 through 18 born between 1978 and 1983. The survey has been used for a wide variety of studies, but grades and GPA are the primary school outcome measures. As Rees and Sabia (2010) note, "the Adolescent Health study did not administer formal achievement tests such as are available in the National Education Longitudinal Survey of 1988." Instead, it contains the Peabody Picture Vocabulary Test (PPVT), a commonly used test of receptive vocabulary. It is described on the NLSY79 website (which used this test along with reading and math achievement assessments) as follows: "The PPVT-R [an updated version of the original] consists of 175 stimulus words and 175 corresponding image plates. Each image plate contains 4 black-and-white drawings, one of which best represents the meaning of the corresponding stimulus word. … Starting in 1998, the administration of the PPVT-R was largely limited to 4- and 5-year-old children".[44]

Reardon makes use of the results of Add Health's PPVT scores for six of the 17 observations used to estimate the growth in the reading gap for cohorts born between 1974 and 2001. But PPVT asks respondents to point at a picture when told a word. No reading is involved.

Add Health provides low estimates of the income-achievement gap in reading for students born between 1978 and 1983 when that gap is alleged to be much lower than in later years. Those lower estimations could easily be due to serious measurement error driven by a test of reading skills that does not require the student to read.

[44] https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/peabody-picture-vocabulary-test-revised [accessed January 15, 2020].

Study of Early Child Care and Youth Development (SECCYD)

Although Reardon says his surveys are nationally representative, SECCYD is not. According to U.S. Department of Health and Human Services (2018), the study recruited ten hospitals willing to participate as recruiting grounds for mothers willing and available to participate in a multi-year study. The sample description explicitly says its "data are not representative in the statistical sense, and therefore inference to the nation as a whole is not possible. Comparisons to other databases, national or otherwise, should be made with extreme caution."[45]

The website for SECCYD identifies use of the Woodcock-Johnson Tests, which are designed to be IQ tests. It is not clear how these might translate into the reading or math achievement tests found in the other surveys. In any event, the survey does not meet the criterion of being nationally representative.

Trends in the SES-Achievement Gap from High-quality Surveys in Reardon (2011b)

With reliance on SES-achievement gap data constructed from different birth cohorts found in large-scale surveys, it is natural to want to include as many different survey data sets as possible. As delineated above, however, six of Reardon's choices do not meet current scientific quality standards. The three early surveys rely upon student reports of household income (Talent, NLS, and HS&B); three later surveys lack reliable test information and/or pertain to non-representative samples (Prospects, Add Health, and SECCYD). These error-prone data points yield gap estimates that are biased downward. Since they tend to relate to early birth cohorts, they distort the trends that are estimated when combined with higher-quality data available later.

We estimate a revised trend line using Reardon's data that restricts observations of gaps to the high-quality surveys in Reardon (2011b). These cover the birth cohorts 1961-2001 (NLSY79 [two cohorts], NELS, NLSY97 [three cohorts], ELS, ECLS-K, and ECLS-B). We supplement these limited observations with an additional observation added from Reardon and Portilla (2016) who update the SES-achievement gap analysis to include the ECLS-K2010.[46] (The

[45] https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/21940/versions/V6/summary [accessed 2/1/2020].

[46] While the three ECLS data sets are included among the high-quality surveys, their outcome data are tests of kindergarten readiness in math and reading. Reardon and Portilla (2016), who specifically analyze trends in these

ECLS-K2010 observation became available subsequent to the publication of Reardon (2011b) and extends the birth cohorts available to 2005).

Appendix Figure A1 replicates the Reardon analysis of estimated 90-10 income-achievement gaps in math and reading for the birth cohorts 1961-2005 using only the high-quality surveys.[47] All of these data come from well-regarded, nationally representative surveys that obtained income information from parents rather than students. One striking fact emerges from this figure: Simple linear regressions that estimate trends in math and reading from the ten observations taken from these surveys show *perfectly flat trends* with no significant change in the income-achievement gap over this time period.

Trends from Two Sets of Psychometrically Linked Surveys

One concern with the overall approach in Reardon (2011b) is that it compares test score gaps (measured in standard deviations) across a large number of structurally different tests. This kind of comparisons of trends across tests can give incorrect inferences even about the direction of change (Ho (2009); Holland (2002)). The problems of such comparisons are particularly severe when the tests have different designs and scales.

It is possible to use two subsets of Reardon's data to investigate the potential impact of testing artifacts on the trend results. The first set of surveys relying on the same testing, NLSY79 and NLSY97, was administered to cohorts born as early as 1961 and as recently as 1981, a twenty-year interval.[48] The second set, ECLS-K, ECLS-B, and ECLS-K2010, provides information on income and achievement for cohorts born as early as 1993 and as late as 2005, a twelve-year interval. Together, these two sets of studies overlap nearly half (28 of 59 years) of the entire time span covered by Reardon (2011b) and over half (15 of 26 years) of the period between 1974 and 2001 when Reardon is more confident of the accuracy of estimated gap growth, identified as at least "30 to 40 percent".

If the income gap had truly increased, it should be evident in these two intertemporally linked surveys. But, at roughly 1.3 s.d., the average income-achievement gap in math and

data, argue that these SES-achievement gaps are comparable to those for later ages in the other surveys because gaps do not change much over the school years.

[47] The point estimates are estimated from observation of the points displayed in Reardon (2011b), Figures 5.1 and 5.2. Reardon and Portilla (2016) provide estimated gaps from the three ECLS surveys.

[48] Household income is measured in the same way in the two surveys, and the non-response rate for NLSY97 is 3 percent, about the same as the 2 percent for NLSY79.

reading estimated by NLSY97 is no different from the one observed twenty years earlier. Over the twelve years between the administration of ECLS-K and ECLS-K2010, the 90-10 math achievement gap, as reported by Reardon and Portilla (2016), declines by 0.13 s.d. and the gap for reading drops by 0.21 s.d.[49]

Neither set of observations reveals any increase in the income-achievement gap. Reardon (2011b) concedes that "the income achievement gap as measured in the NLSY97 cohort is virtually identical to the gap in the NLSY79 cohort, born twenty years earlier" (p. 96). But, he says, "the NLSY cohort was born in the early 1980s, just as the trend" upward is about "to begin" once more. This survey was administered too soon to discern "a rising gap among the 1980s and 1990s cohorts." The ECLS data, analyzed in Reardon and Portilla (2016), removes uncertainty about whether or not the NLSY comparison was simply due to observations outside the relevant range.

Conclusion

Both estimates from all high-quality surveys (Appendix Figure A1) and those from psychometrically matched surveys indicate no substantial change in the income-achievement gap in math and reading over time. The size of the gap remains constant at roughly the same level as the SES-achievement gap reported in this paper, indicating that analyses based on different surveys and different underlying SES measures actually come to the same broad conclusion. In other words, the apparently contradictory results between our analysis of SES-achievement trends and those of Reardon (2011b) are completely resolved by recognition of the systematic measurement error embedded in specific surveys used in Reardon.

[49] The math gap declines from 1.3 s.d. to 1.17 s.d.; the reading gap from 1.26 s.d. to 1.05 s.d.

Appendix C: Extrapolating Achievement in the Tails of the SES Distribution

As discussed in section 4.2, it is not always possible with the available survey data to create an SES index that reliably distinguishes achievement differences for the tails of the SES distribution. Our main analysis includes data from a given test regime-age-subject-year survey if the SES distribution provides observations of students within the top and bottom quartile of the constructed SES distribution. If such observations are available, it is possible to obtain reliable information on average achievement in the top and bottom quartiles.

An alternative approach, applied by Reardon (2011b), Chmielewski and Reardon (2016), and Chmielewski (2019) (see section 2.2), is to extrapolate achievement from the observed range to points farther in the tail of the distribution. It is just the shape of the SES-achievement relationship in the tails of the SES distribution that is the object of the overall analysis. We do not use this extrapolation approach because it necessarily involves making strong assumptions about the shape of the SES-achievement distribution that cannot be verified by the data.

The inherent difficulties with this extrapolation can be illustrated with our SES-achievement data. Figure 8 in the text shows the SES histogram for Main-NAEP in 1990, incorporating student data on parental education and home-possession items. The top panel of Appendix Figure A10 presents this SES distribution in terms of the achievement distribution corresponding to the SES percentiles that are observed. The bars in the figure indicate the percentile ranges of the SES distribution covered by the categorical elements of the SES distribution along with the average achievement in the category. For example, the highest SES category identified in the data covers the top 26.5 percent of the distribution and has an average achievement of 278 scale points.

The prior literature has used an extrapolation approach in order to estimate achievement gaps in the unobserved extreme tails of the SES distribution, such as the 90-10 SES-achievement gap. In the example of the Main-NAEP 1990, the challenge—and the objective of the extrapolation—is to estimate the achievement of somebody at the 90^{th} percentile of the SES distribution, because it is not possible to observe any achievement differences of students in the top 26.5 percent of the distribution. The method applied in the prior literature has two steps. In the first, it is assumed that a person at the midpoint of each percentile range has achievement equal to the category average (see the data points added to each bar in the figure). This step involves imposing linearity on the distribution within each category. In the second step, a

regression function (linear or cubic) is estimated from all of the data points, and this function is evaluated at the 90th percentile. This step assumes that the pattern of achievement farther down the SES distribution provides a good reflection of the SES-achievement distribution in the tails of the SES distribution.

Appendix Figure A10 shows the fitted linear and cubic regressions following this procedure. Even accepting the underlying distributional assumptions in the two steps, the estimated achievement of a person at the 90th percentile would differ between the linear and the cubic extrapolation by 0.14 s.d., even though the two alternative methods are very close in the midranges of the SES distribution.[50]

The extrapolation uncertainty becomes particularly severe when the range of possible categories of the SES distribution observed in the tails is limited. The limited information in the tails is generally exacerbated when the SES distribution is constructed from a single variable—for example, categorical parental income in Reardon (2011b) or parental education in Chmielewski (2019). The bottom panel of Appendix Figure A10 shows the SES-achievement distribution for PISA 2015 when just parental education is used to identify the SES distribution (instead of the combined information on parental education and home-possession items in Figure 8). Fully 46 percent of the students are found in the top parental-education category, and 89 percent are in the top three categories. This is the information based on which achievement at the 90th percentile is extrapolated in Chmielewski (2019)'s analysis of the PISA data.

The problem of choosing a functional form for extrapolation is not very important with so few categorical observations, but the problem of what SES percentile corresponds to the average achievement in the broad categories is crucial. The single SES measure cannot adequately resolve this issue, but the choice can obviously lead to enormous differences in the extrapolated achievement levels and thus in the estimated SES-achievement gaps. In fact, the top education category averages over 40 percent 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-achievement gaps based solely on surveyed parental education.

[50] When estimating 90-10 gaps, Reardon (2011b) and Chmielewski (2019) use a mixture of cubic projection modified to linear projection when there are greater than 20 percent of observations in the top or bottom category. This corresponds to whether the assumed data point in the respective category from the first step falls outside the top or bottom 10 percent of the overall distribution.

Table A1. Surveys and Subjects 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

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

Table A2. Data for Trend Analyses by Survey, Test Year, Age, and Subject

					Overall achievement dispersion		SES-achievement gap
Survey	Test year	Age	Subject	Mean	90-10	75-25	75-25
pisa	2000	15	math	493	2.622	1.364	1.309
pisa	2003	15	math	482	2.564	1.329	1.074
pisa	2006	15	math	474	2.394	1.293	1.024
pisa	2009	15	math	488	2.429	1.302	1.011
pisa	2012	15	math	481	2.371	1.285	0.949
pisa	2015	15	math	470	2.386	1.262	0.904
pisa	2000	15	reading	504	2.632	1.372	1.182
pisa	2003	15	reading	494	2.495	1.332	1.047
pisa	2009	15	reading	500	2.423	1.304	0.979
pisa	2012	15	reading	497	2.237	1.214	0.842
pisa	2015	15	reading	498	2.497	1.347	0.695
pisa	2000	15	science	499	2.525	1.409	1.200
pisa	2003	15	science	490	2.560	1.410	1.085
pisa	2006	15	science	489	2.733	1.522	1.133
pisa	2009	15	science	502	2.487	1.355	1.026
pisa	2012	15	science	497	2.349	1.298	0.939
pisa	2015	15	science	496	2.502	1.384	0.852
timss	1995	14	math	487	2.577	1.413	0.752
timss	1999	14	math	502	2.463	1.327	0.821
timss	2003	14	math	504	2.314	1.222	0.894
timss	2007	14	math	509	2.213	1.190	0.860
timss	2011	14	math	509	2.199	1.154	0.781
timss	2015	14	math	518	2.406	1.286	0.898
timss	1995	14	science	521	2.588	1.365	0.651
timss	1999	14	science	515	2.379	1.257	0.839
timss	2003	14	science	527	1.991	1.048	0.835
timss	2007	14	science	520	2.002	1.074	0.843
timss	2011	14	science	524	1.982	1.053	0.763
timss	2015	14	science	529	2.007	1.059	0.773
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	0.961
naep	2000	14	math	276	2.817	1.475	
naep	2005	14	math	279	2.774	1.454	1.142
naep	2007	14	math	281	2.742	1.443	1.134
naep	2009	14	math	283	2.776	1.456	1.158
naep	2011	14	math	284	2.767	1.458	1.157
naep	2013	14	math	285	2.771	1.460	1.227
naep	2015	14	math	282	2.832	1.496	1.268
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	0.877
naep	2005	14	reading	262	2.351	1.219	0.943
naep	2007	14	reading	263	2.303	1.177	0.934
naep	2009	14	reading	264	2.267	1.160	0.922
naep	2011	14	reading	265	2.269	1.172	0.955
naep	2013	14	reading	268	2.280	1.180	1.013
naep	2015	14	reading	265	2.321	1.188	1.010

(continued on next page)

Table A2. (continued)

Survey	Test year	Age	Subject	Mean	Overall achievement dispersion 90-10	Overall achievement dispersion 75-25	SES-achievement gap 75-25
naepltt	1978	13	math	264	2.846	1.532	0.985
naepltt	1982	13	math	268	2.442	1.298	0.761
naepltt	1986	13	math	269	2.276	1.183	0.782
naepltt	1990	13	math	270	2.314	1.208	0.759
naepltt	1992	13	math	273	2.273	1.202	0.792
naepltt	1994	13	math	275	2.372	1.236	0.854
naepltt	1996	13	math	275	2.334	1.209	0.839
naepltt	1999	13	math	275	2.415	1.256	
naepltt	2004	13	math	281	2.399	1.247	0.830
naepltt	2008	13	math	281	2.470	1.253	0.857
naepltt	2012	13	math	285	2.535	1.318	0.933
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	0.576
naepltt	1990	13	reading	257	2.029	1.050	0.798
naepltt	1992	13	reading	260	2.218	1.162	0.981
naepltt	1994	13	reading	258	2.224	1.141	0.923
naepltt	1996	13	reading	258	2.228	1.135	0.969
naepltt	1999	13	reading	259	2.156	1.146	0.966
naepltt	2004	13	reading	259	2.054	1.070	0.815
naepltt	2008	13	reading	260	2.072	1.045	1.080
naepltt	2012	13	reading	263	2.067	1.062	1.013
naepltt	1973	14	math	266			
naepltt	1973	17	math	304			
naepltt	1978	17	math	300	2.580	1.408	1.020
naepltt	1982	17	math	298	2.425	1.305	0.886
naepltt	1986	17	math	302	2.305	1.251	0.945
naepltt	1990	17	math	305	2.314	1.287	0.900
naepltt	1992	17	math	307	2.215	1.196	0.820
naepltt	1994	17	math	306	2.278	1.178	0.932
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	0.878
naepltt	2008	17	math	306	2.174	1.164	0.865
naepltt	2012	17	math	306	2.246	1.169	0.960
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	0.921
naepltt	1988	17	reading	290	2.110	1.115	0.644
naepltt	1990	17	reading	290	2.298	1.202	0.691
naepltt	1992	17	reading	290	2.429	1.232	0.790
naepltt	1994	17	reading	288	2.466	1.290	0.789
naepltt	1996	17	reading	287	2.339	1.220	0.783
naepltt	1999	17	reading	288	2.349	1.215	0.809
naepltt	2004	17	reading	285	2.445	1.234	0.855
naepltt	2008	17	reading	286	2.484	1.268	0.956
naepltt	2012	17	reading	287	2.372	1.229	0.987

Notes: Mean: average student achievement. Overall achievement dispersion: achievement difference between 90th and 10th (75th and 25th) percentiles of achievement distribution. SES-achievement gap: achievement difference between top and bottom quartiles of SES distribution. See Figures 1, 2, and 5 for additional information.

Table A3. Racial SES Distribution over Time

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

Notes: Data source: LTT-NAEP.

Table A4. Components of Background Surveys, Main-NAEP 1990 and PISA 2015

	Main-NAEP 1990	PISA 2015
Highest education of parents	1 = Did not finish high school 2 = Graduated high school 3 = Some education after high school 4 = Graduated college	0 = None 1 = Grade 6 2 = Grade 9 3 = < High school 4 = High school graduation 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?	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 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?

Notes: Main-NAEP 1990 and PISA 2015 chosen for expositional purposes only as examples with a low and high number of information categories.

Table A5. Acronyms for Sources of Data Analyzed by Reardon (2011b)

Acronym	Survey
Talent	Project Talent
NLS	National Longitudinal Study of 1972
HS&B	High School and Beyond
NLSY79	National Longitudinal Study of Youth 1979
NELS	National Education Longitudinal Study
Prospects	The Congressionally Mandated Study of Educational Growth and Opportunity
Add Health	National Longitudinal Study of Adolescent to Adult Health
NLSY97	National Longitudinal Study of Youth 1997
ELS	Education Longitudinal Study
SECCYD	Study of Early Child Care and Youth Development
ECLS-K	Early Childhood Longitudinal Study-Kindergarten, 1997
ECLS-B	Early Childhood Longitudinal Study, Birth Cohort

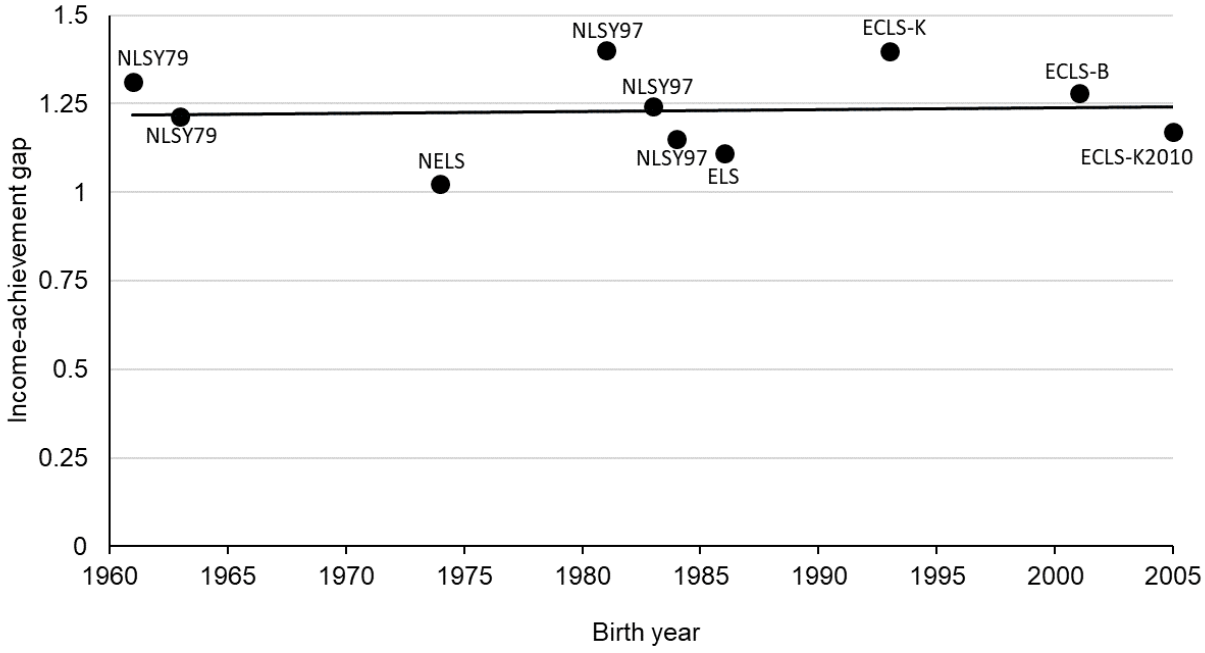
Table A6. Correlation between SES Index and Family Income in ELS

	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

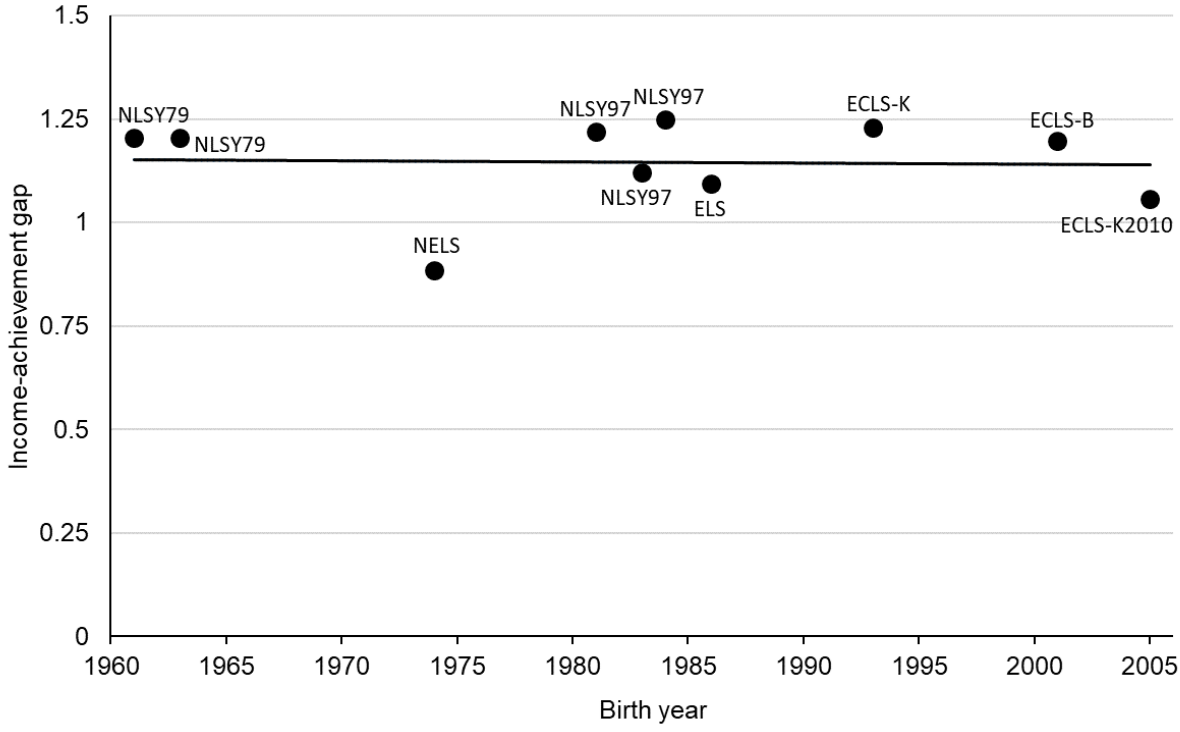
Notes: Data source: 1988 Education Longitudinal Study (ELS).

Figure A1. Estimated 90-10 Income-Achievement Gaps from High-quality Surveys, Birth Cohorts 1961-2005

Panel A: Math

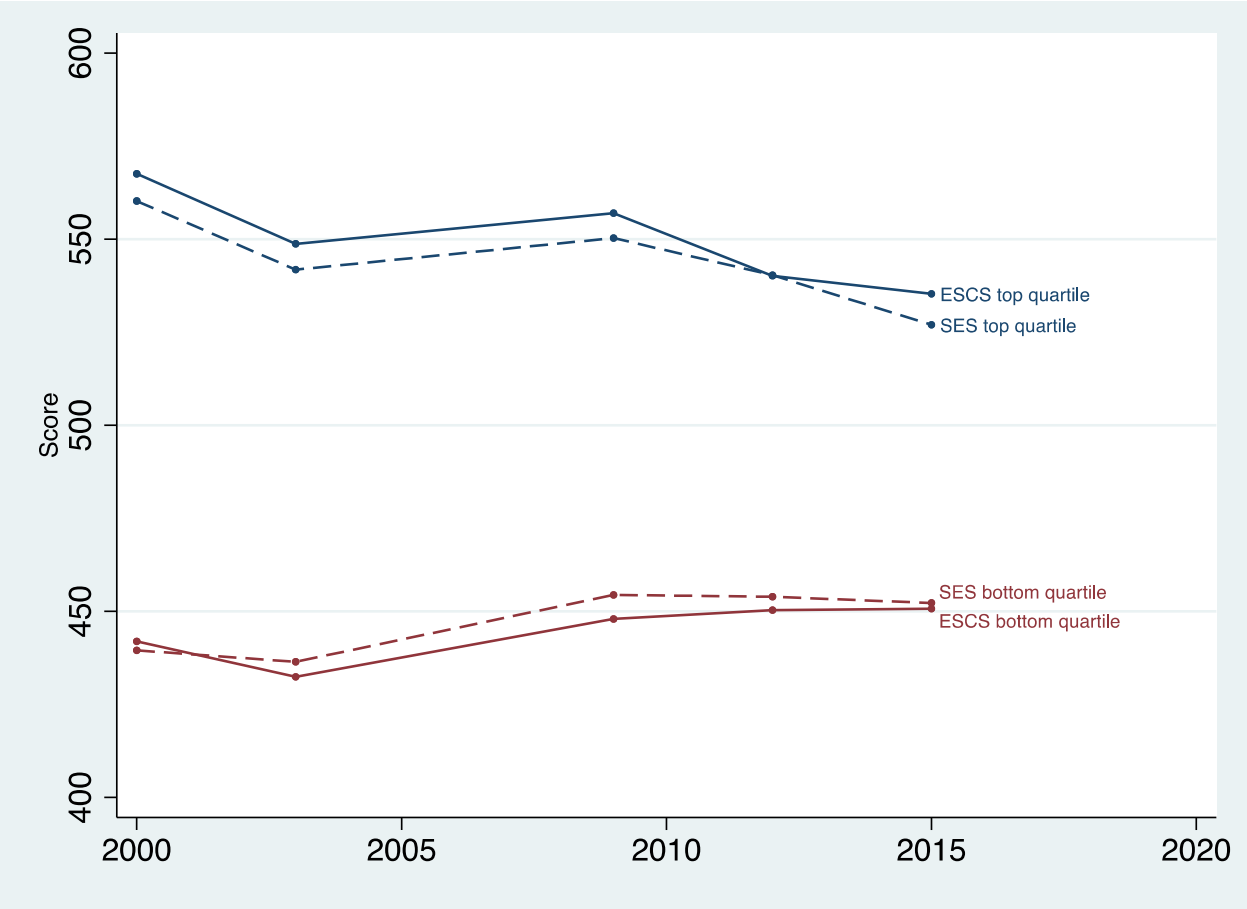


Panel B: Reading



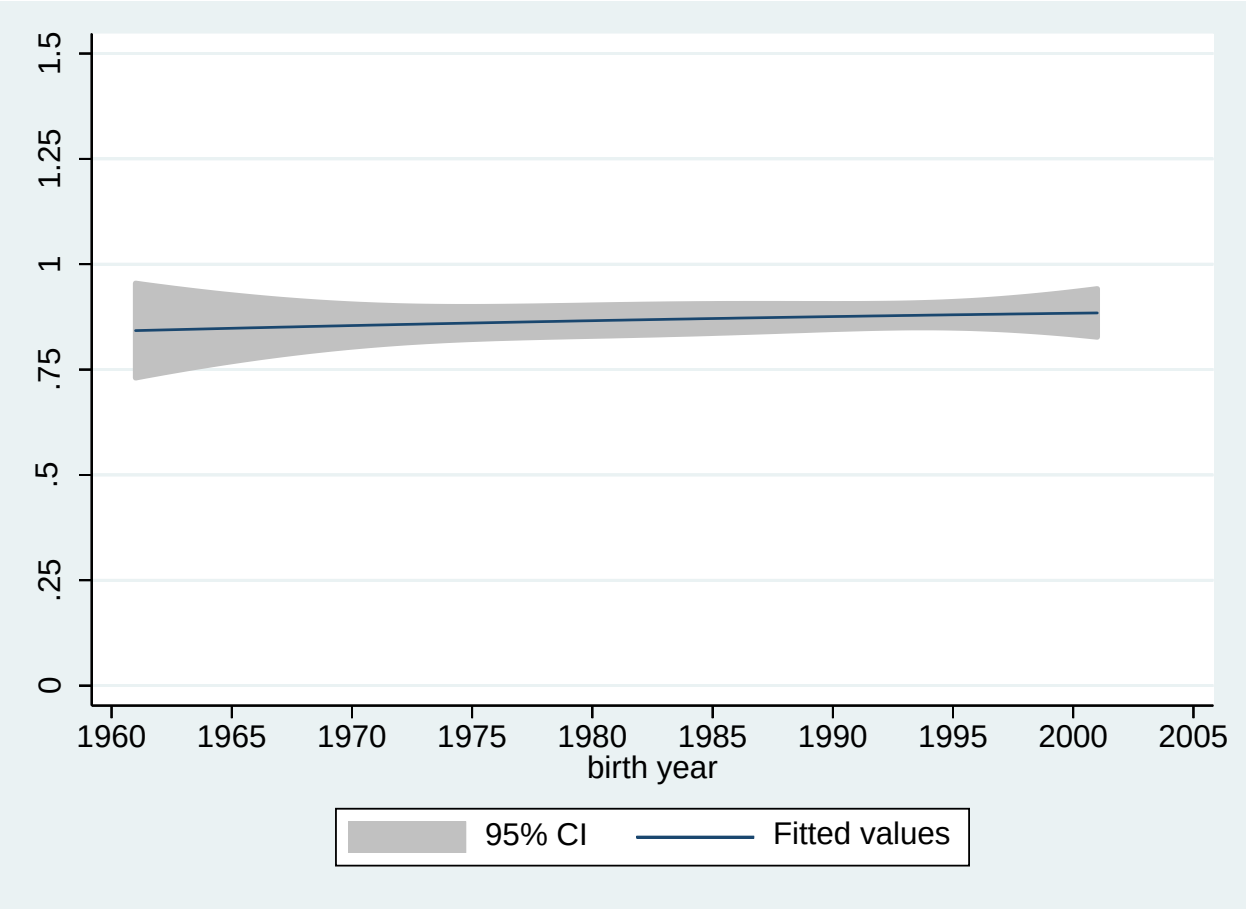
Notes: Replication of the Reardon (2011b) analysis of the 90-10 income-achievement gap using only high-quality surveys, birth cohorts 1961-2005. See Table A5 for acronyms.

Figure A2. Achievement Trends of the Top and Bottom Quartile in PISA based on PISA's ESCS Index and our SES Index by Test Year



Notes: 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.

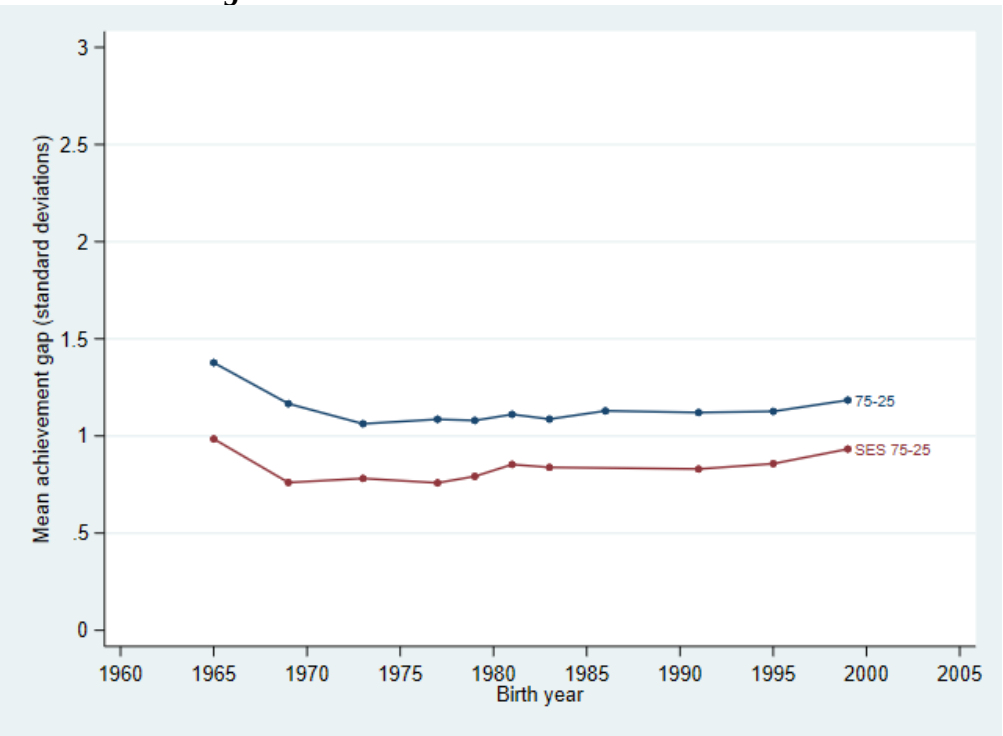
Figure A3. Trend in the SES-Achievement Gap with Confidence Interval, Birth Cohorts 1961-2001



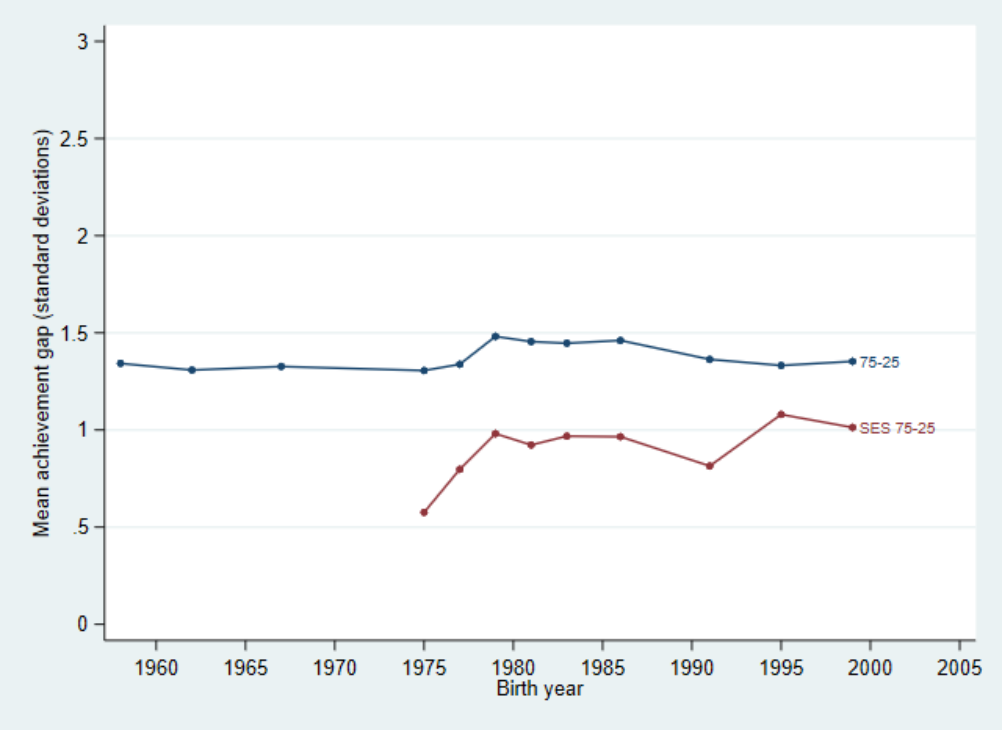
Notes: Achievement difference between the students in the top and bottom quartiles of the SES distribution (75-25 SES-achievement gap). See Figure 2 for details.

Figure A4. LTT-NAEP SES-Achievement Gaps and Overall Achievement Dispersion

Panel A: Math Age 13



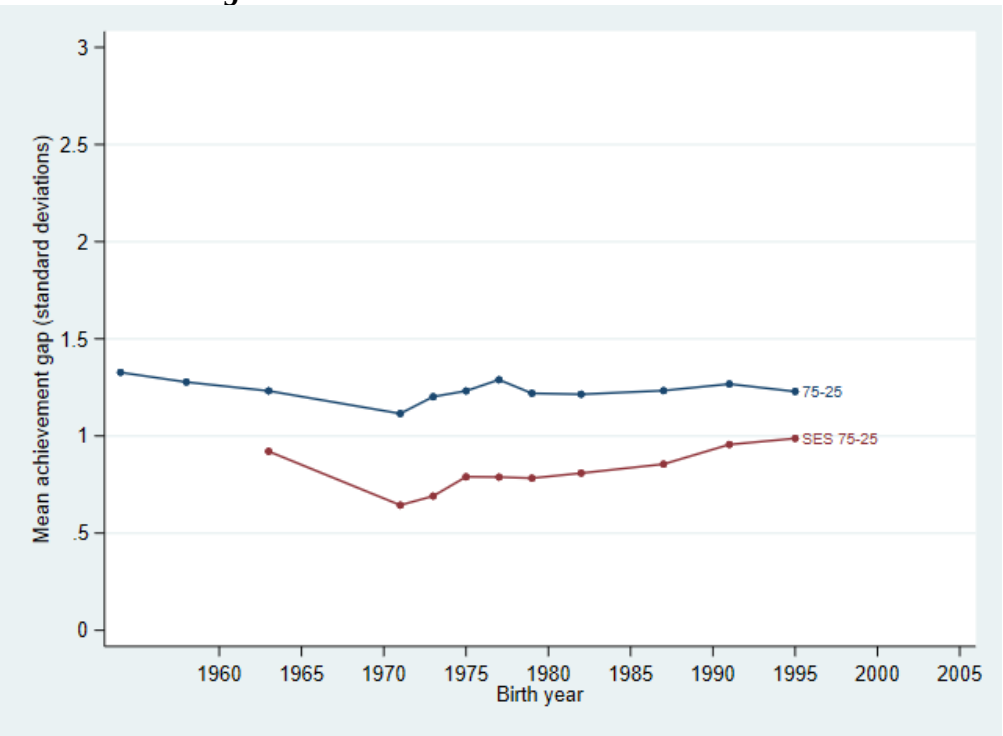
Panel B: Reading Age 13



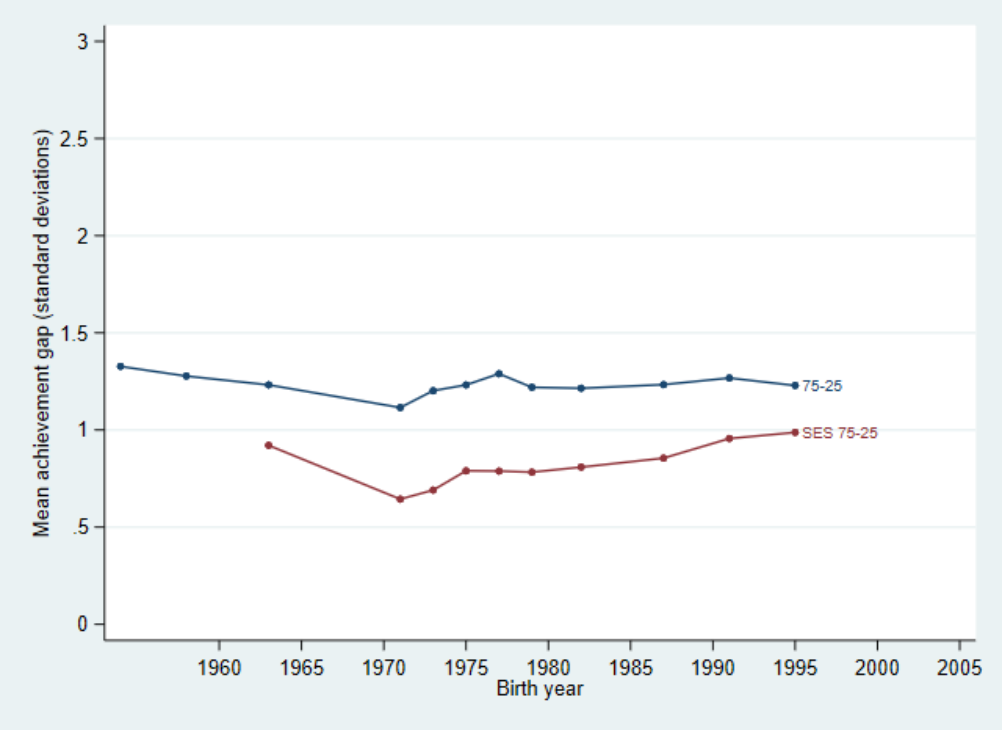
(continued on next page)

Figure A4. (continued)

Panel A: Math Age 17



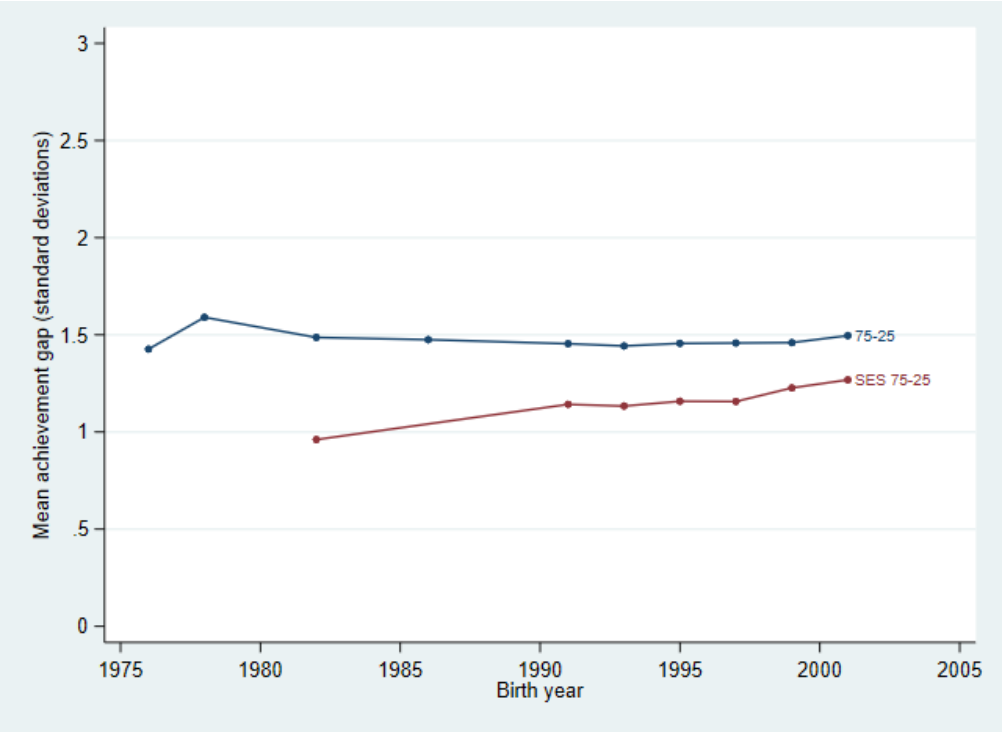
Panel B: Reading Age 17



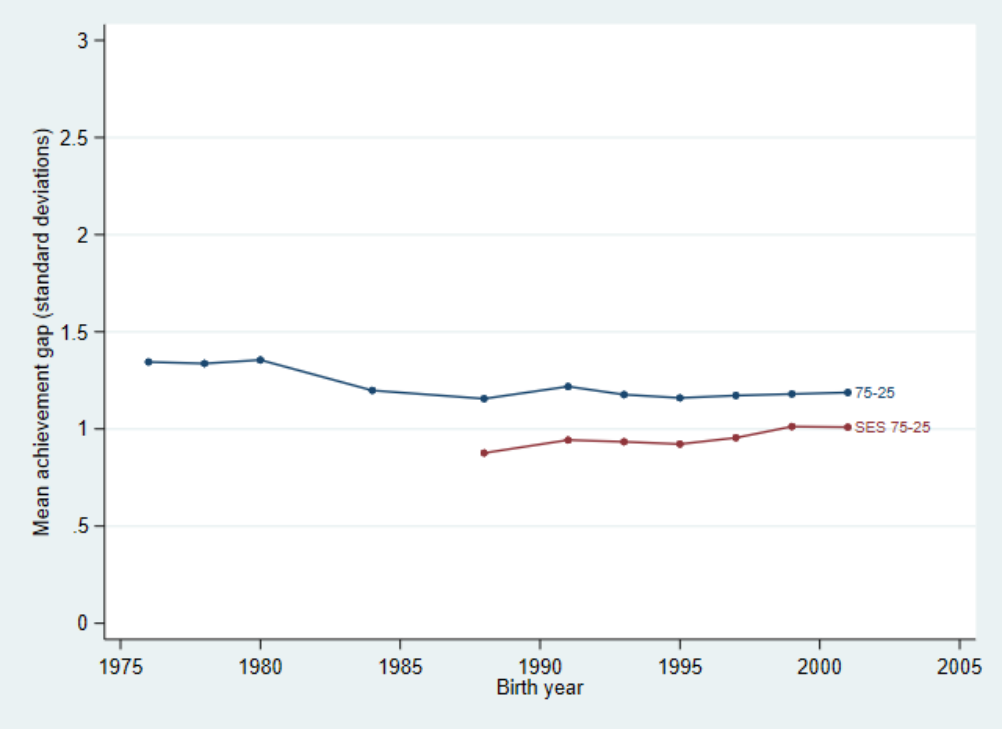
Notes: SES 75-25: achievement difference between the students in the top and bottom quartiles of the SES distribution. 75-25: overall achievement difference between the students at the 75th and 25th percentiles of the achievement distribution. Normalized achievement is measured in standard deviations (of the test closest to 2000).

Figure A5. Main-NAEP SES-Achievement Gaps and Overall Achievement Dispersion

Panel A: Math



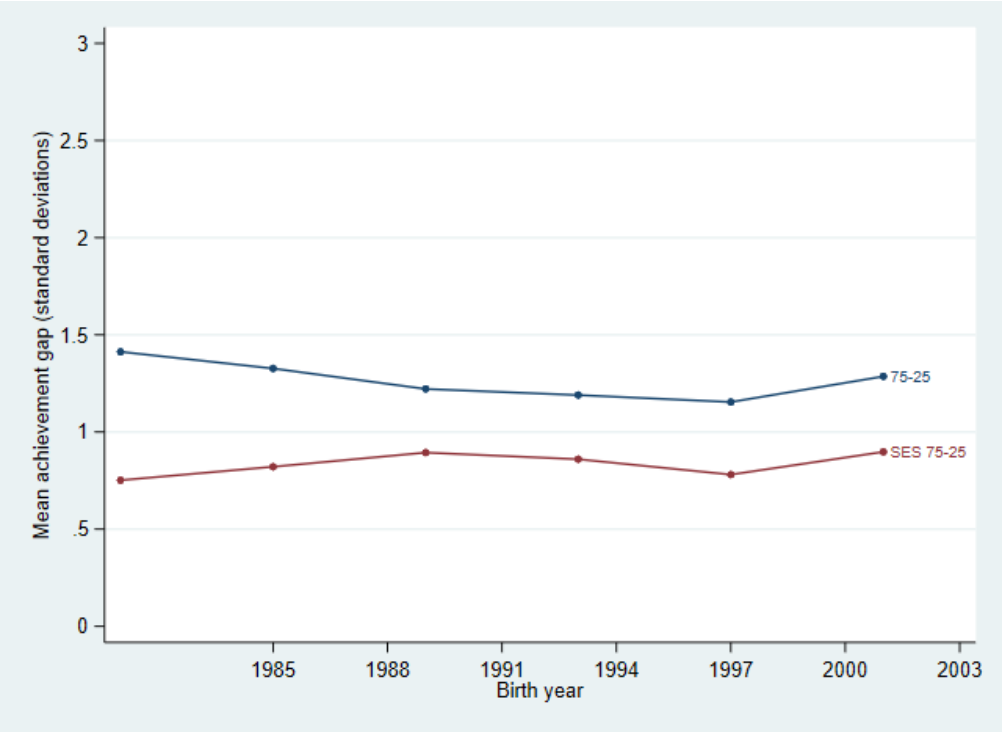
Panel B: Reading



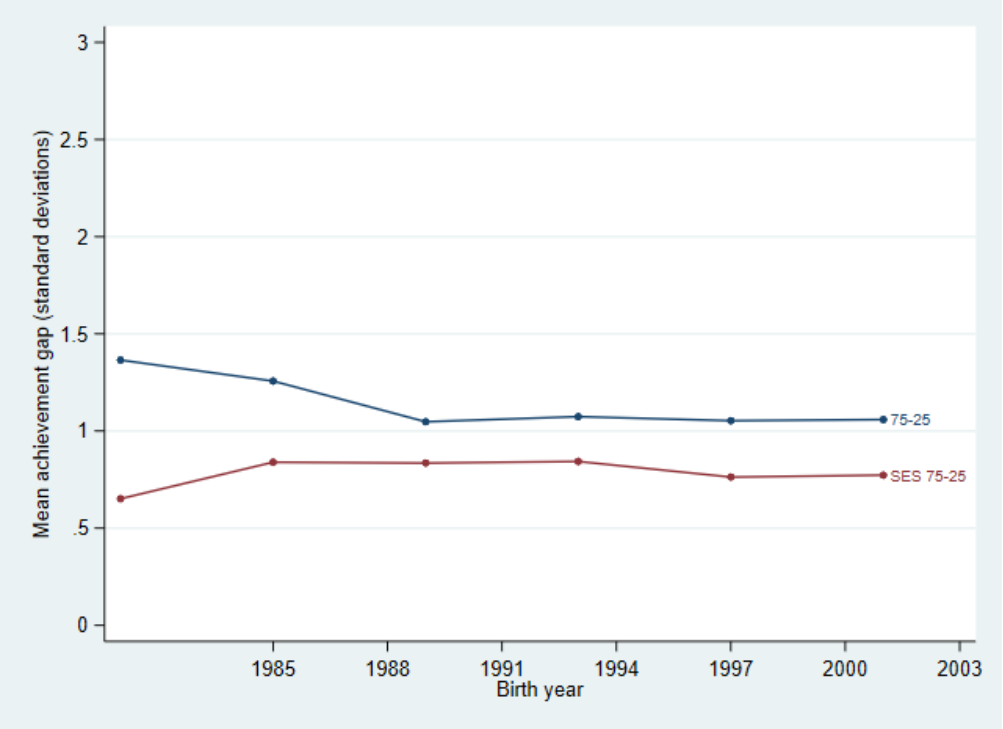
Notes: SES 75-25: achievement difference between the students in the top and bottom quartiles of the SES distribution. 75-25: overall achievement difference between the students at the 75th and 25th percentiles of the achievement distribution. Normalized achievement is measured in standard deviations (of the test closest to 2000).

Figure A6. TIMSS SES-Achievement Gaps and Overall Achievement Dispersion

Panel A: Math



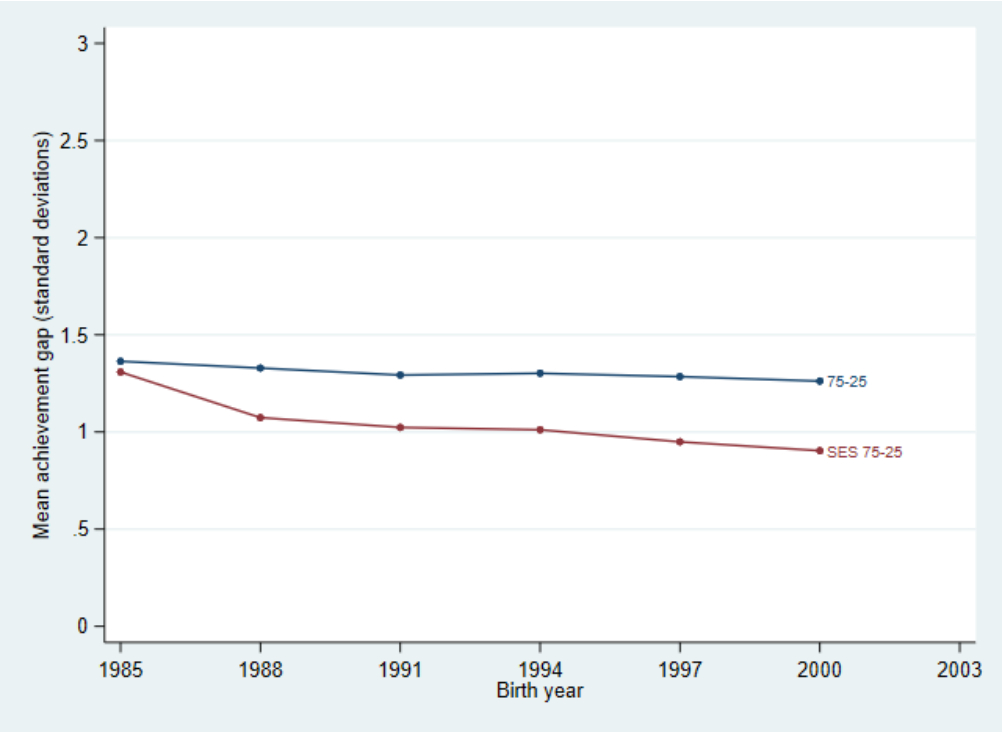
Panel B: Science



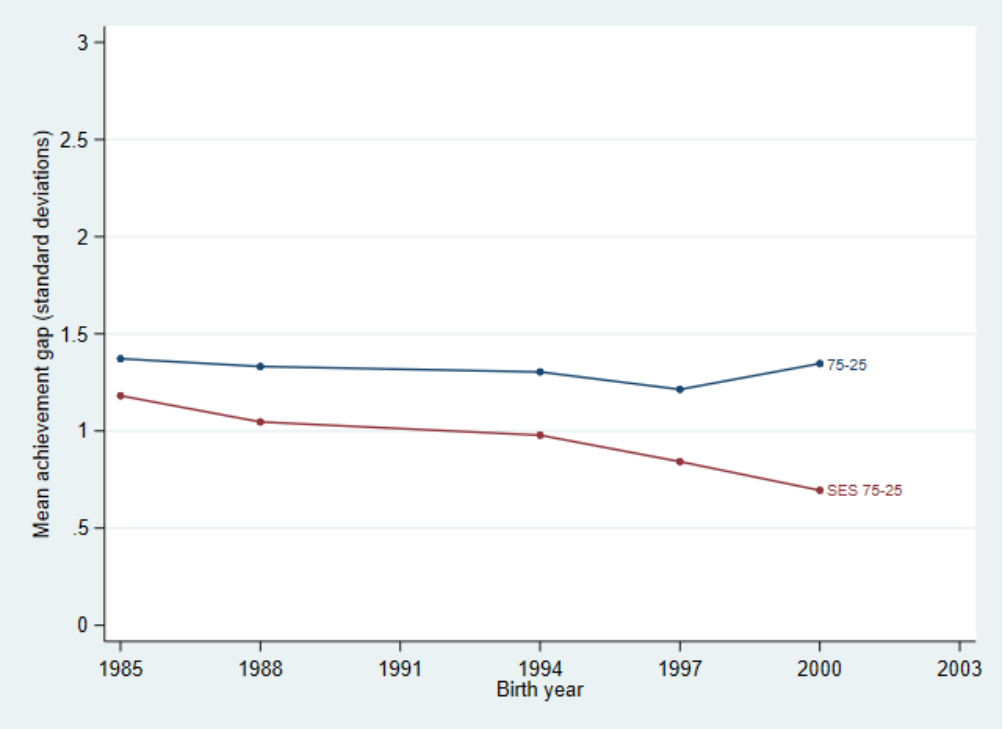
Notes: SES 75-25: achievement difference between the students in the top and bottom quartiles of the SES distribution. 75-25: overall achievement difference between the students at the 75th and 25th percentiles of the achievement distribution. Normalized achievement is measured in standard deviations (of the test closest to 2000).

Figure A7. PISA SES-Achievement Gaps and Overall Achievement Dispersion

Panel A: Math



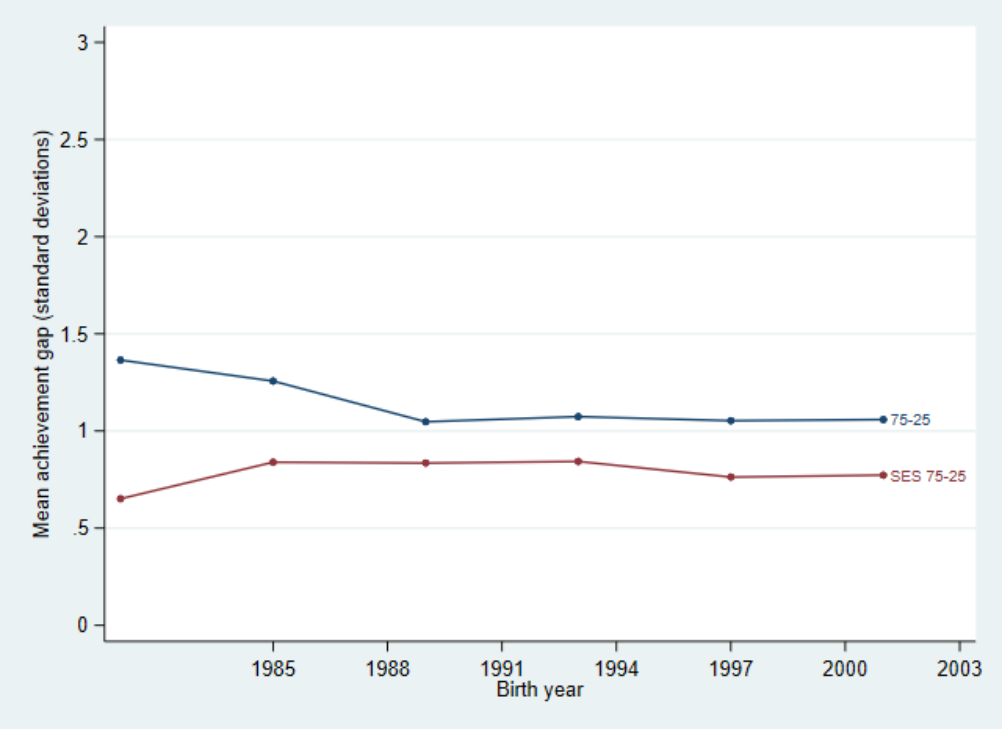
Panel B: Reading



(continued on next page)

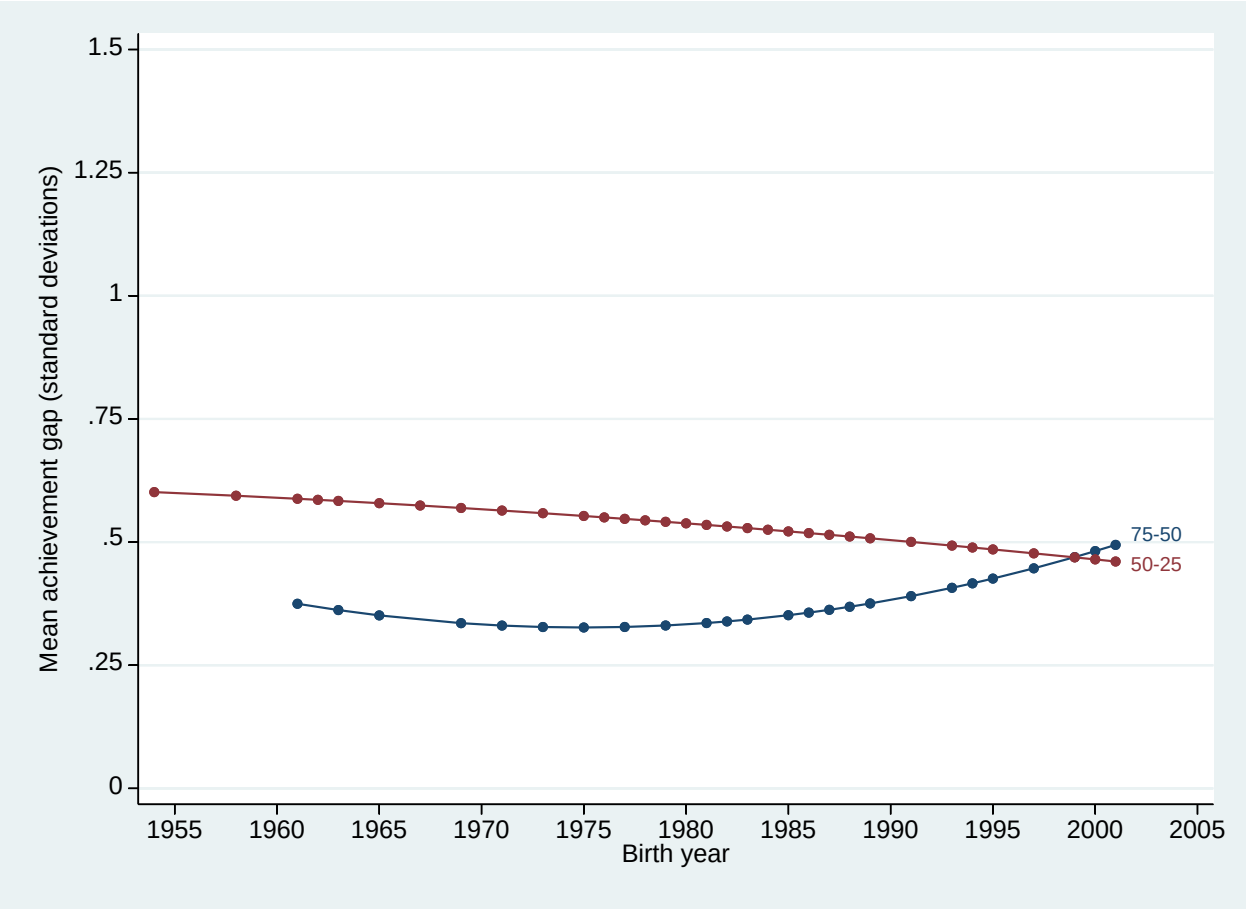
Figure A7. (continued)

Panel C: Science



Notes: SES 75-25: achievement difference between the students in the top and bottom quartiles of the SES distribution. 75-25: overall achievement difference between the students at the 75th and 25th percentiles of the achievement distribution. Normalized achievement is measured in standard deviations (of the test closest to 2000).

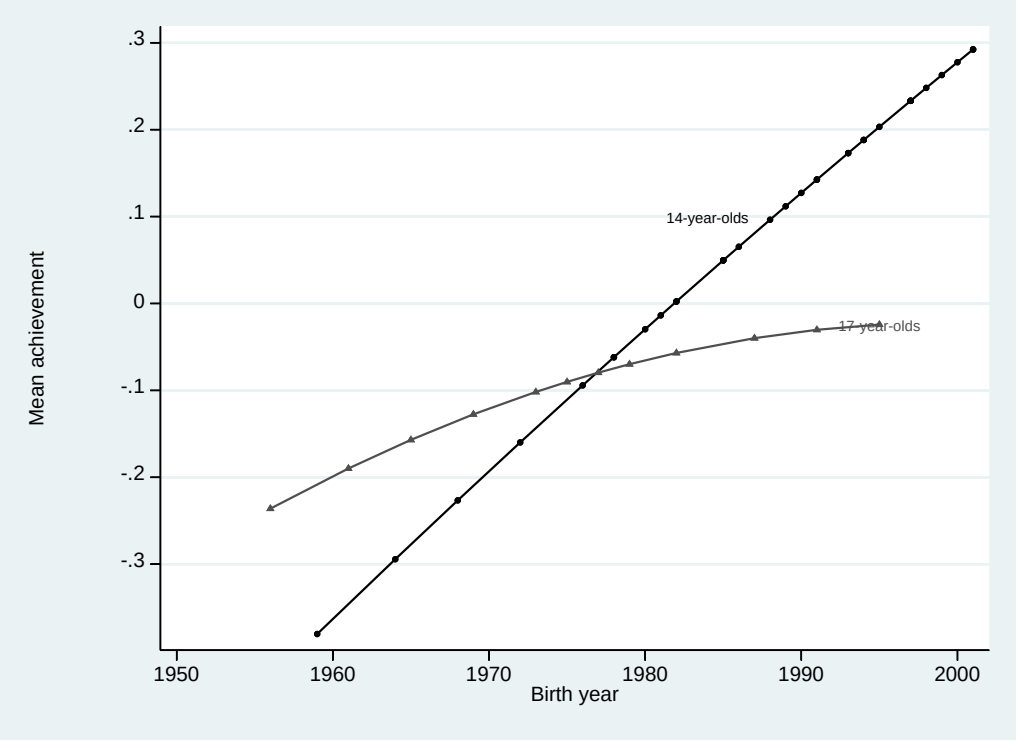
Figure A8. Alternative Calculation of 75-50 and 50-25 SES-Achievement Gaps, Birth Cohorts 1954-2001



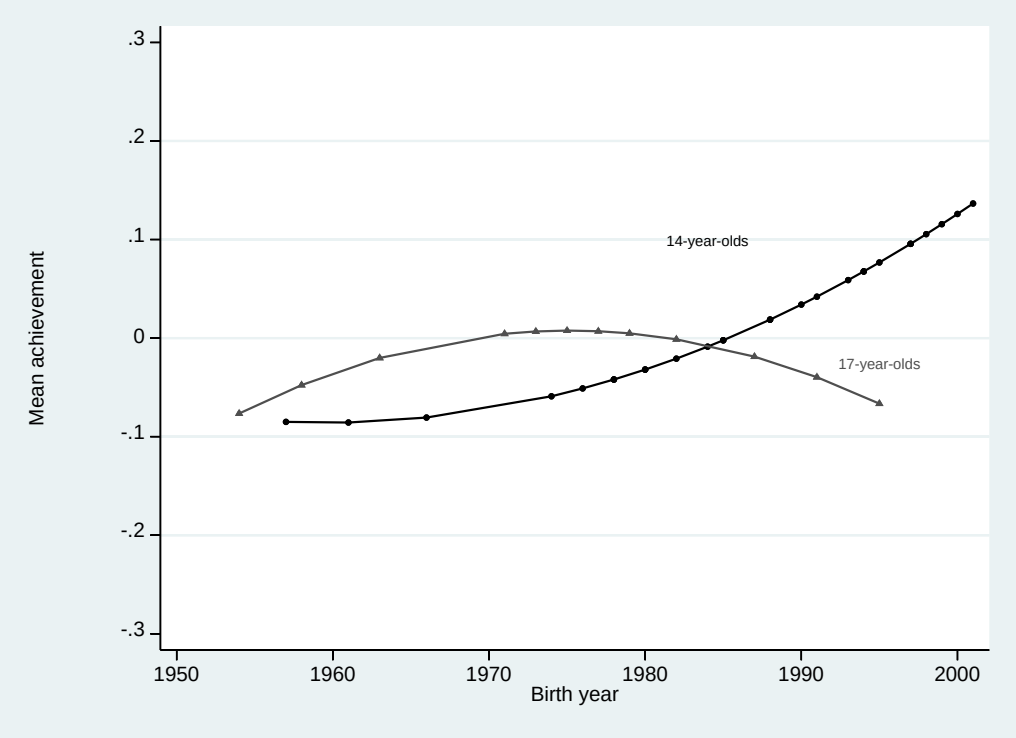
Notes: 75-50 SES-achievement gap: achievement difference between the students in the top quartile and in the 45-55 percentile band of the SES distribution. 50-25 SES-achievement gap: achievement difference between the students in the 45-55 percentile band and the bottom quartile of the SES distribution. See Figure 2 for data and methods.

Figure A9. Achievement Levels of Younger and Older Students by Subject, Birth Cohorts 1954-2001

Panel A: Math



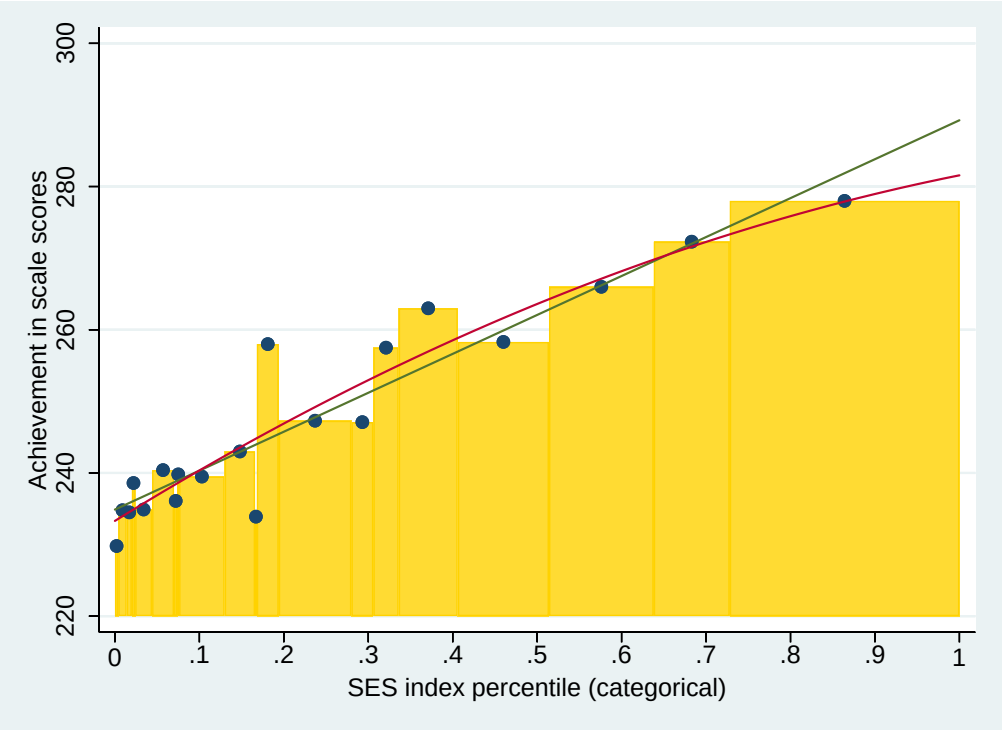
Panel B: Reading



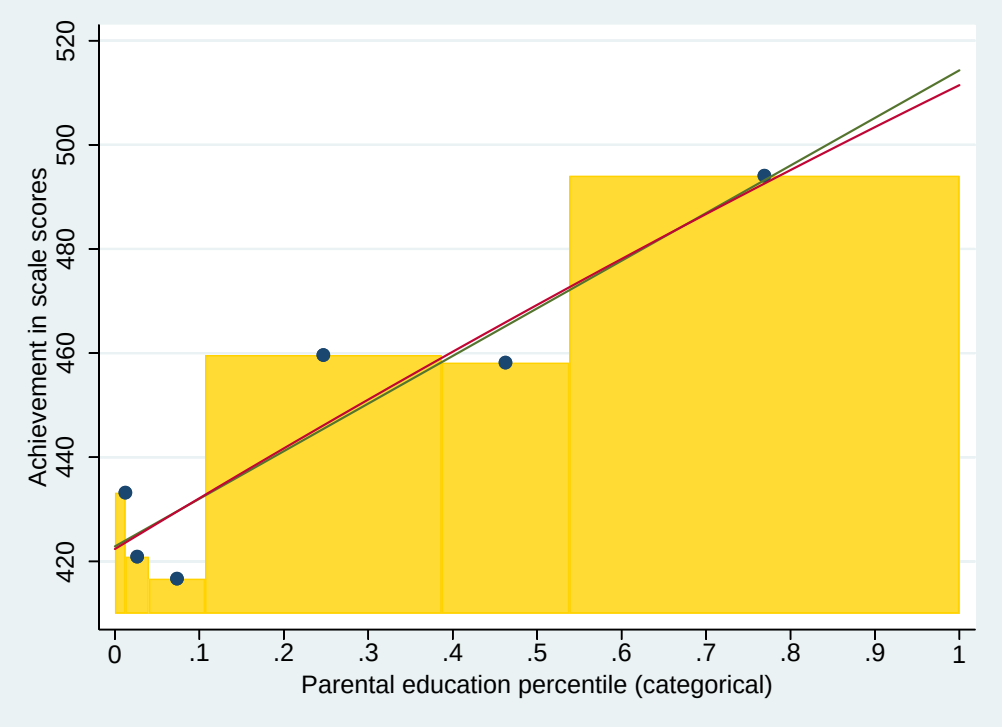
Notes: Average student achievement. See Figures 1 and 5 for data and methods.

Figure A10. Illustration of Regression Prediction for Point Estimation Approach

Panel A: Main-NAEP 1990, SES Index



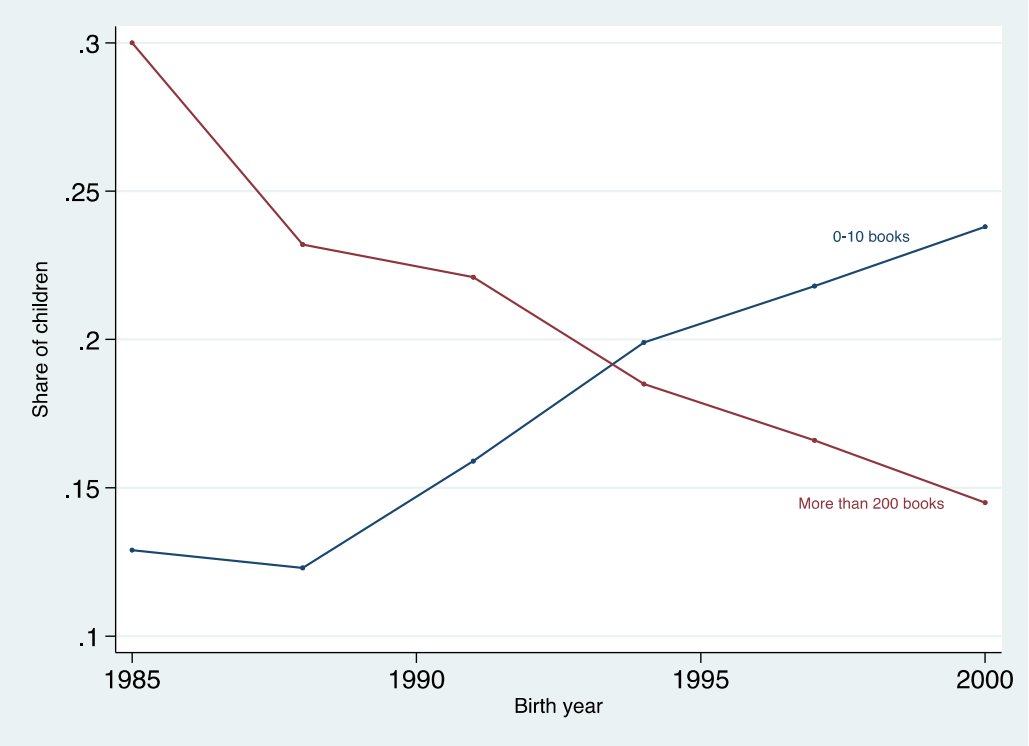
Panel B: PISA 2015, Parental Education



Notes: Average scores by SES category (percentiles of the discrete values of the respective underlying SES measure) with respective category midpoints and linear and cubic regression predictions.

Figure A11. 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

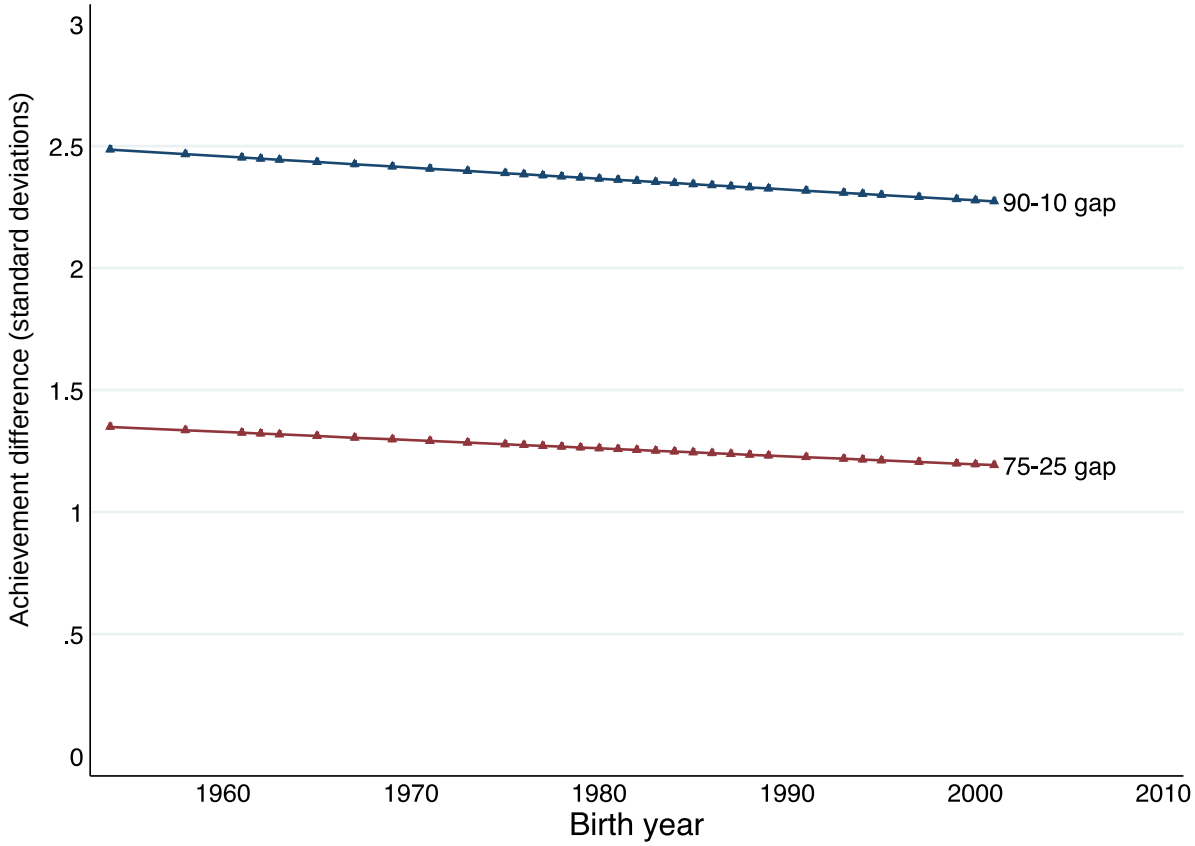
Notes: U.S. student population in PISA.

Table 1. Description of Achievement Data

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

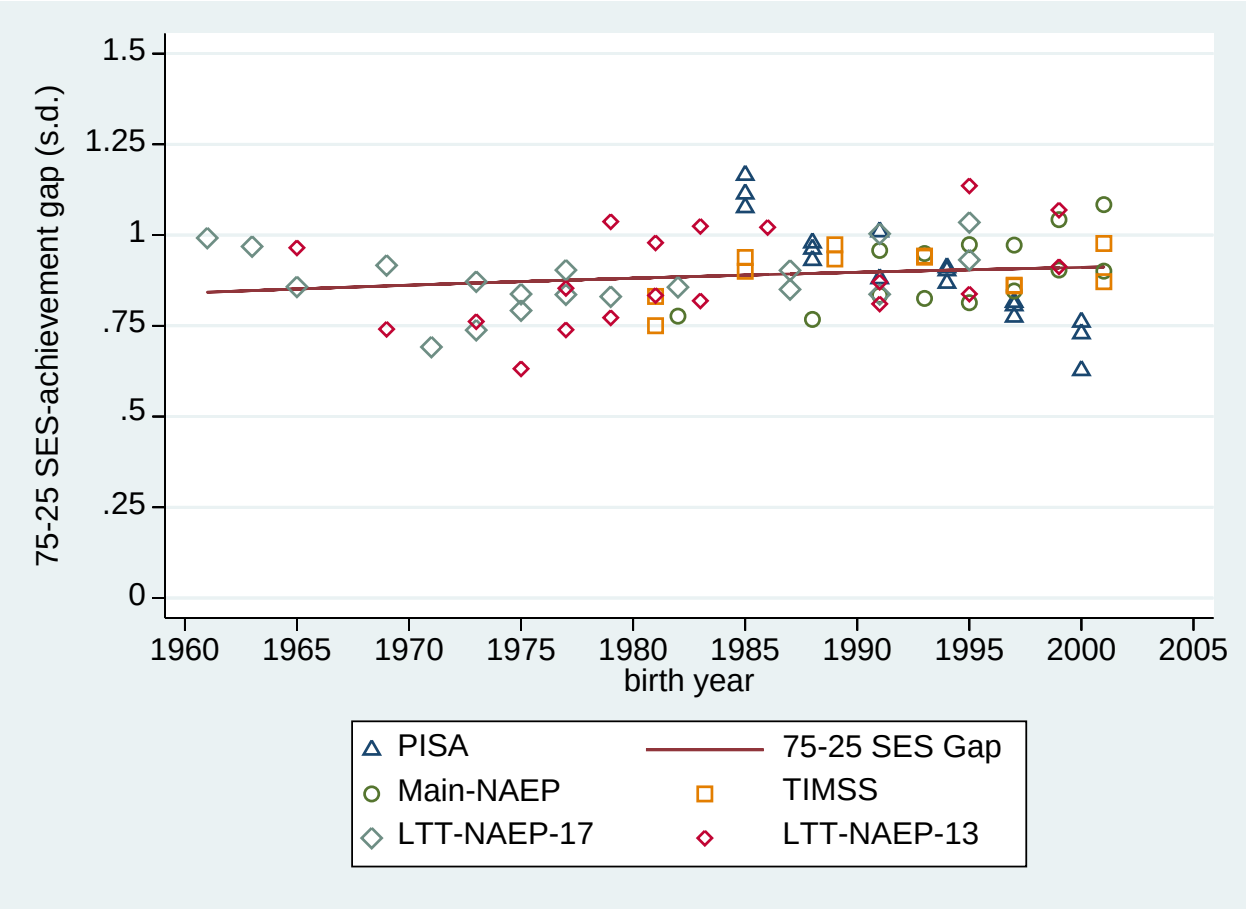
Notes: LTT-NAEP math is first tested in 1973, as opposed to reading which starts in 1971. For the 1973 math, data are only available for achievement levels and not for achievement gaps. Sample sizes for the restricted-use NAEP data are rounded to the nearest 10.

Figure 1. Overall Achievement Dispersion among U.S. Students, Birth Cohorts 1954-2001



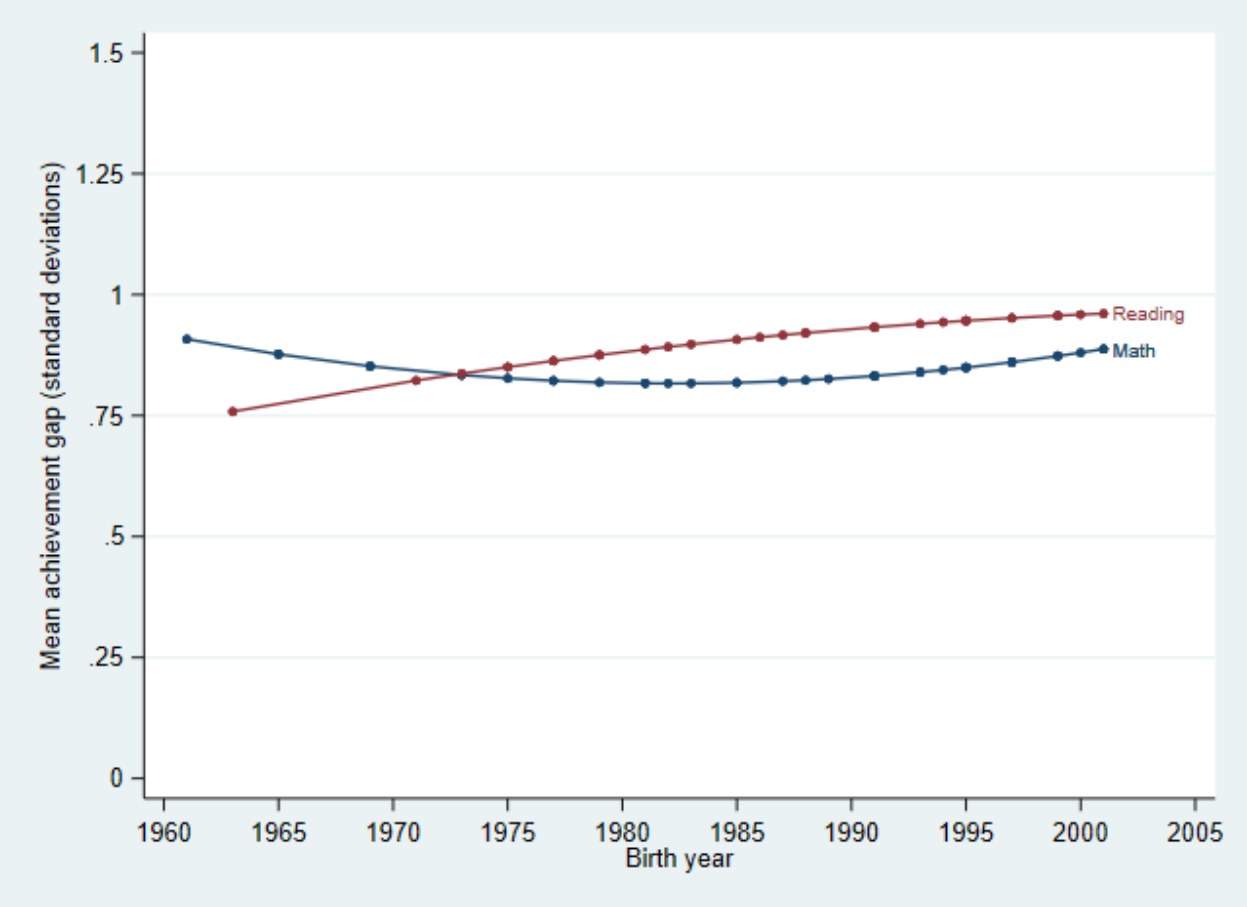
Notes: 75-25 (90-10): overall achievement difference between the students at the 75th and 25th (90th and 10th) percentiles of the achievement distribution. All tests administered by LTT-NAEP, Main-NAEP, PISA, and TIMSS. 1954-2001 birth cohorts, all subjects, all students. Normalized achievement is measured in standard deviations (of the installment of the respective test series closest to 2000). Each marker indicates years where there are one or more underlying observations.

Figure 2. Trend in the SES-Achievement Gap with Underlying Test Data, Birth Cohorts 1961-2001



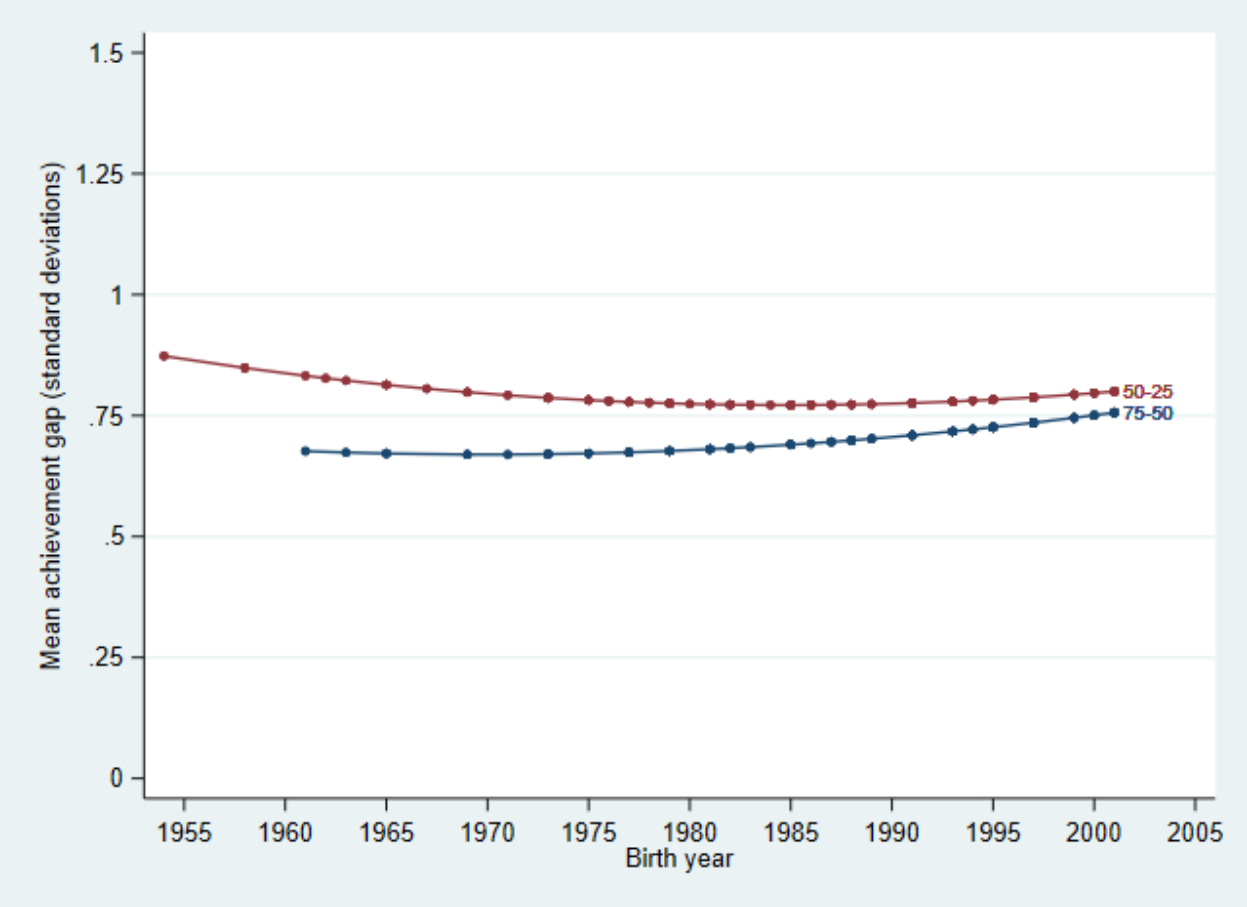
Notes: Achievement difference between the students in the top and bottom quartiles of the SES distribution (75-25 SES-achievement gap). All tests administered by LTT-NAEP, Main-NAEP, PISA, and TIMSS. 1961-2001 birth cohorts, all subjects, all students. Normalized achievement is measured in standard deviations (of the installment of the respective test series closest to 2000). Each marker indicates one organization-subject-age observation. Test data points are adjusted by the fixed effects estimated for equation 2. The trend line for the 75-25 SES-achievement gap is the fitted quadratic from equation 2. The linear and quadratic terms are individually and jointly insignificantly different from zero (individually, $p>0.90$ for each; jointly, $F(2,72)=0.53$, $p=0.59$).

Figure 3. SES-Achievement Gap by Subject, Birth Cohorts 1961-2001



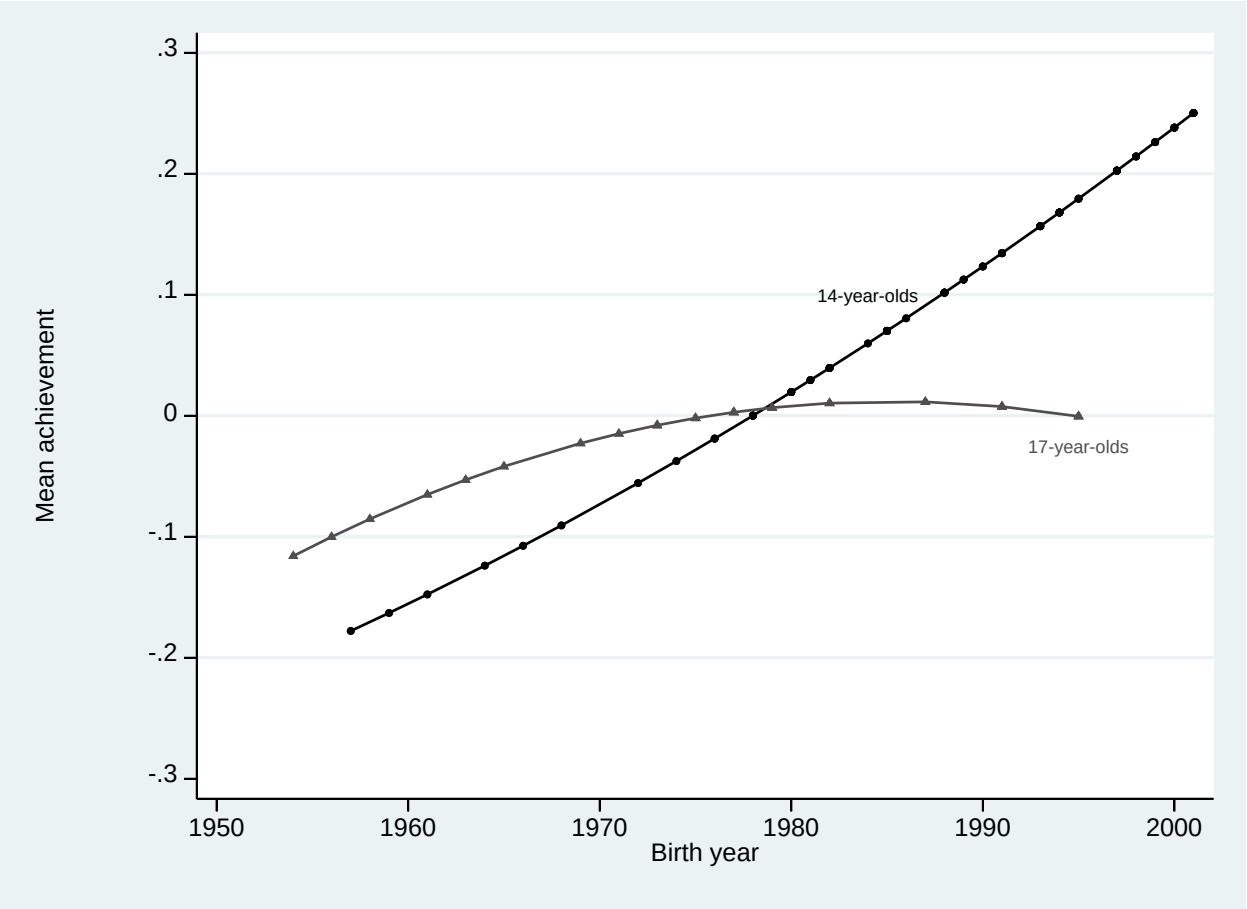
Notes: Achievement difference between the students in the top and bottom quartiles of the SES distribution (75-25 SES-achievement gap), by subject. The markers on the trend lines indicate years where there is reliable information about the SES gaps. See Figure 2 for data and methods.

Figure 4. 75-50 and 50-25 SES-Achievement Gaps, Birth Cohorts 1954-2001



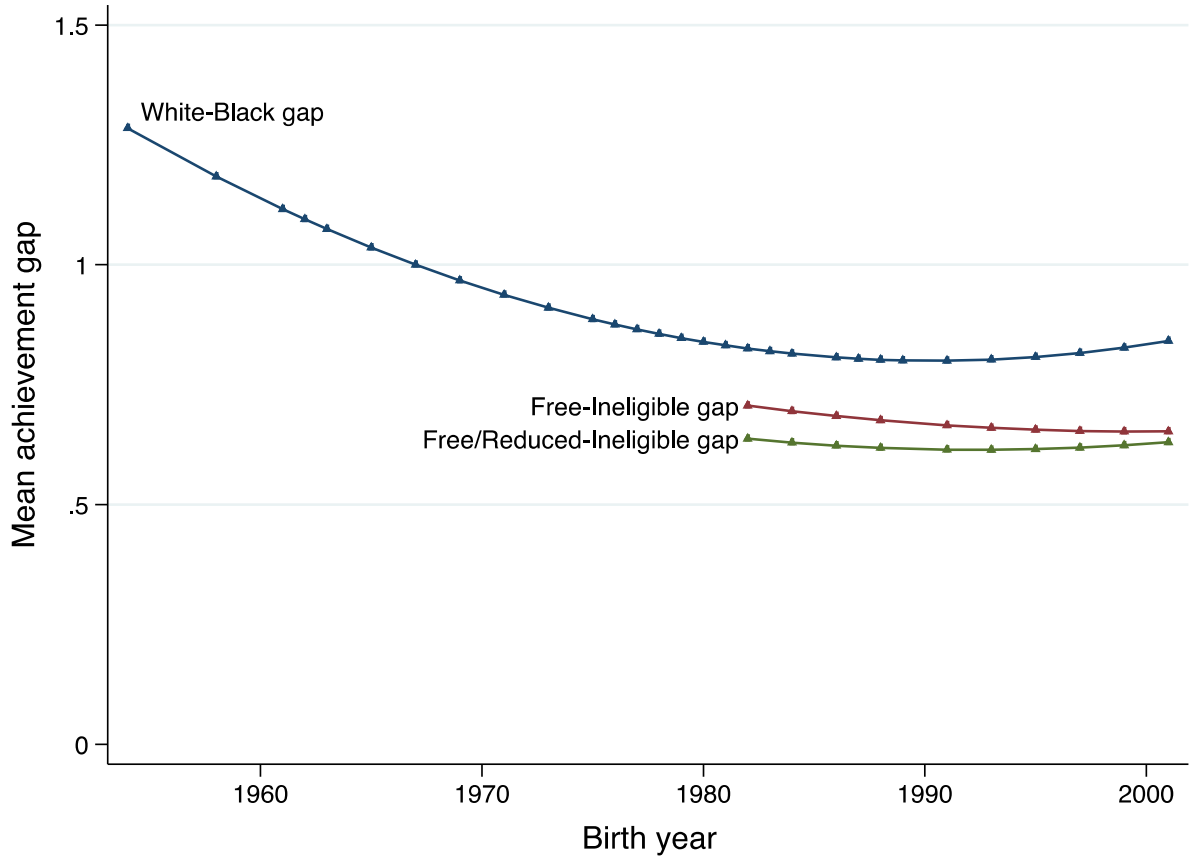
Notes: 75-50 SES-achievement gap: achievement difference between the students in the top quartile and the bottom half of the SES distribution. 50-25 SES-achievement gap: achievement difference between the students in the top half and the bottom quartile of the SES distribution. See Figure 2 for data and methods.

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



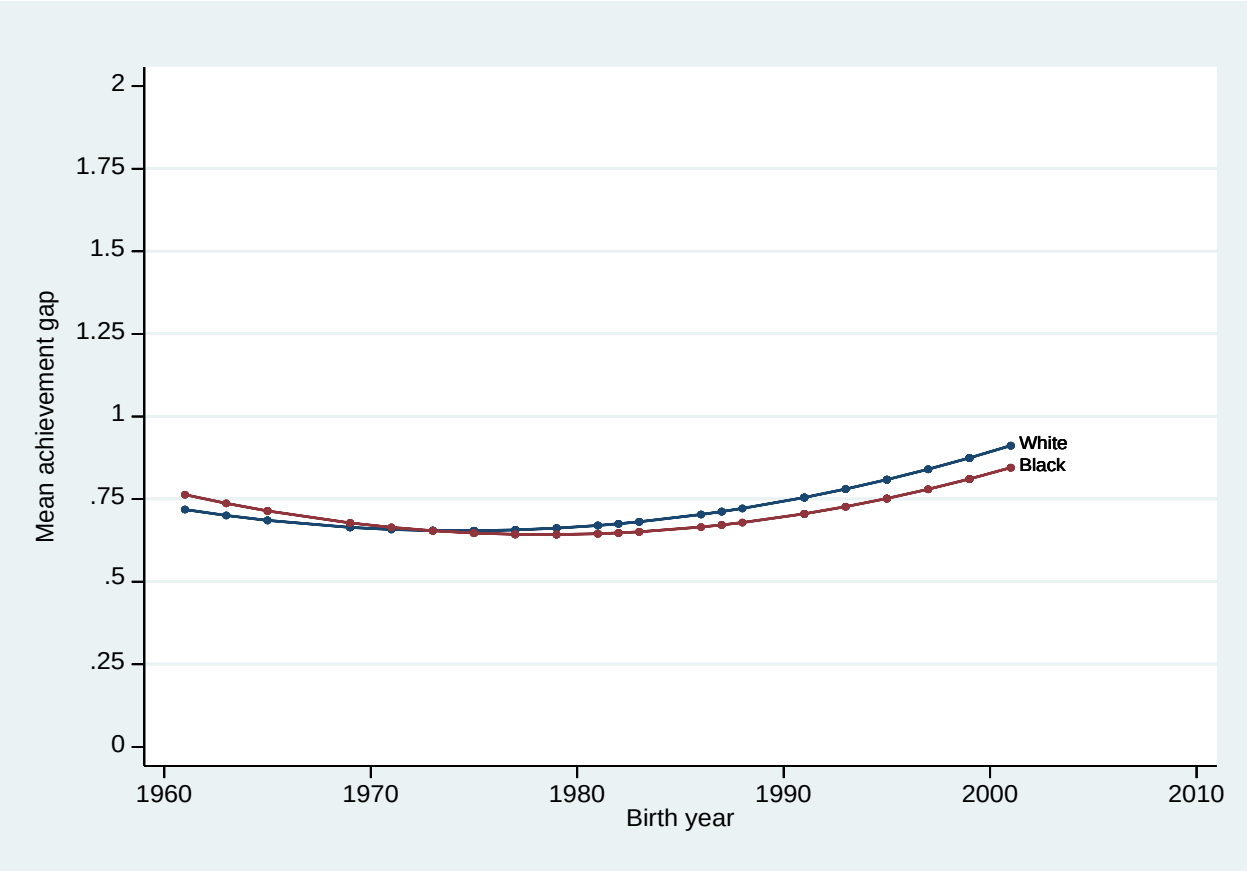
Notes: Average student achievement. Sample: 1954-2001 birth cohorts, all surveys, all subjects, all students. Younger students are those between ages 13 and 15 or in 8th grade, depending on the test. For expositional purposes, younger students are referred to as 14-year-olds. Older students are those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

Figure 6. Achievement Gaps for Subsidized Lunch Eligibility and Black-White Achievement Gaps, Birth Cohorts 1954-2001



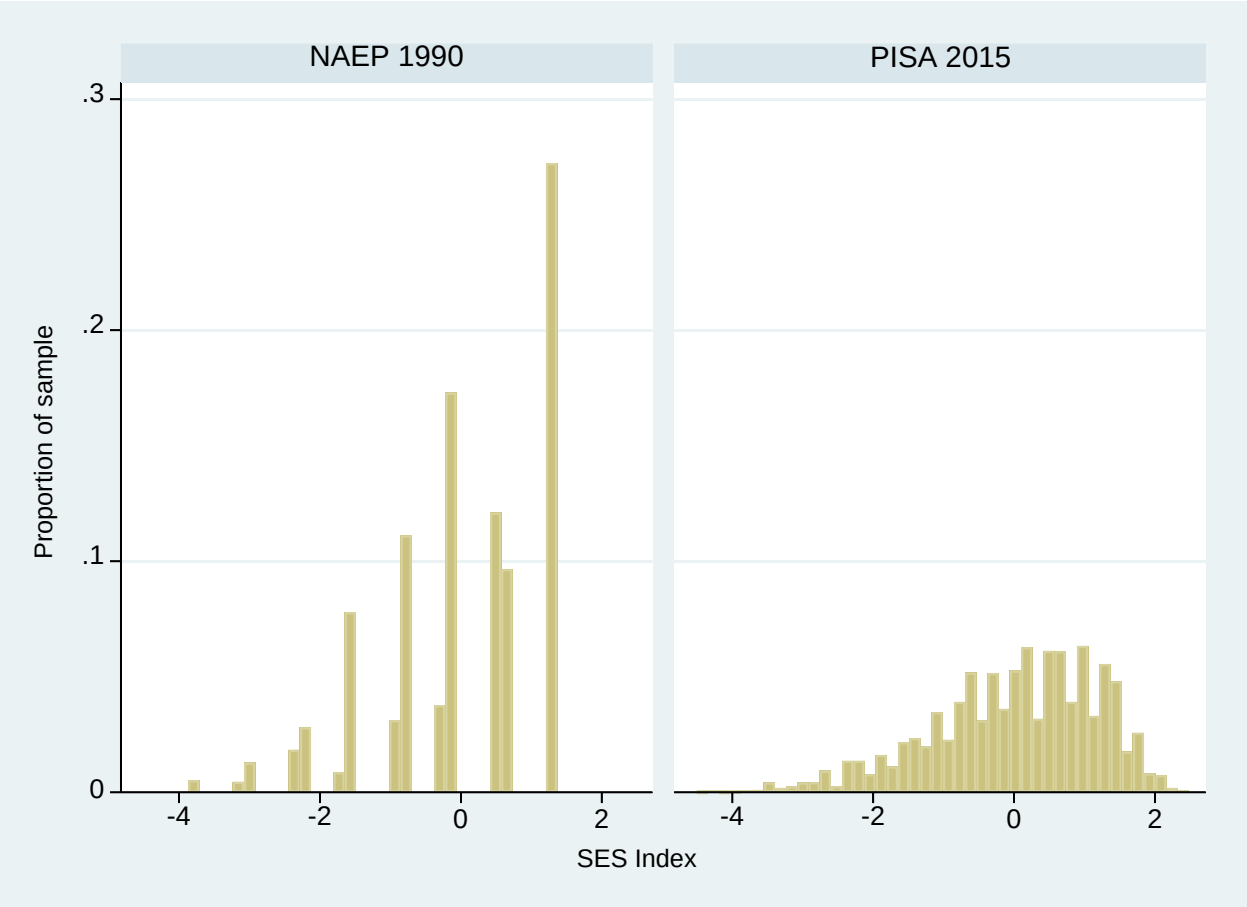
Notes: Samples: For free and reduced-price lunch, 1982-2001 birth cohorts, Main-NAEP surveys, math and reading, all students; for white-black gap, 1954-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, black and white students. See Figure 2 for data and methods. Data on free and reduced-price lunch eligibility are only available for Main-NAEP tests, starting with the 1982 birth cohort.

Figure 7. SES-Achievement Gaps among White and Black Students Separately, Birth Cohorts 1961-2001



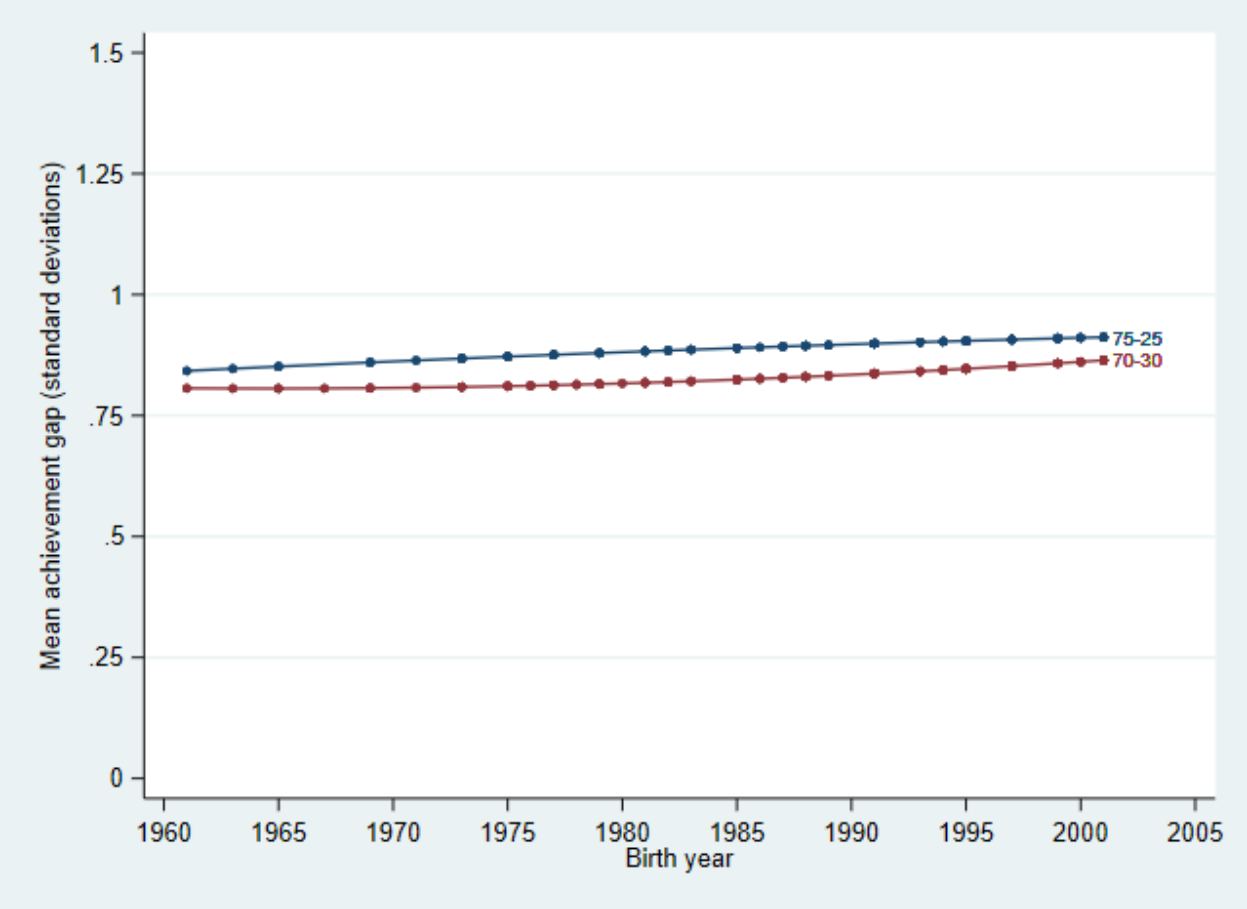
Notes: Achievement difference between the students in the top and bottom quartiles of the SES distribution (75-25 SES-achievement gap), by race. Sample: 1961-2001 birth cohorts, LTT-NAEP and Main-NAEP surveys, math and reading, white and black students. See Figure 2 for data and methods with national (common) SES distribution.

Figure 8. Histograms of SES Distributions, Main-NAEP 1990 and PISA 2015



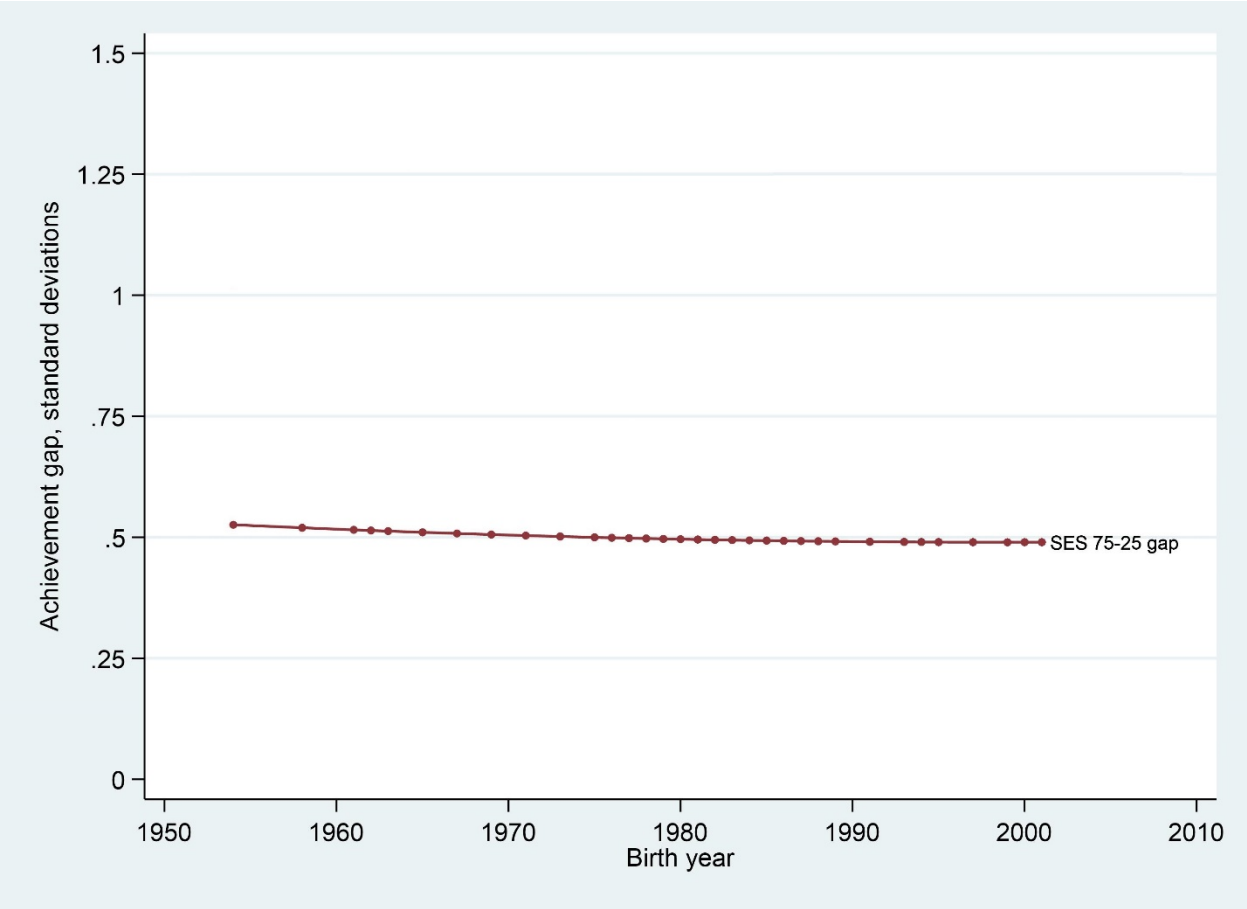
Notes: Frequency of specific values for the Main-NAEP 1990 and PISA 2015 SES distributions calculated by the first principal component of the parental-education and home-possession variables; see text.

Figure 9. 75-25 and 70-30 SES-Achievement Gaps, Birth Cohorts 1961-2001



Notes: Achievement difference between the students in the top and bottom 25 (30) percent of the SES distribution. The markers on the trend lines indicate years where there is reliable information about the SES gaps. See Figure 2 for data and methods. The 75-25 gap trend line is the same as shown in Figure 2. The 70-30 gap trend line is estimated from 91 observations; the joint significance test for the quadratic parameters is $F(2, 82)=1.05$, $p>0.35$.

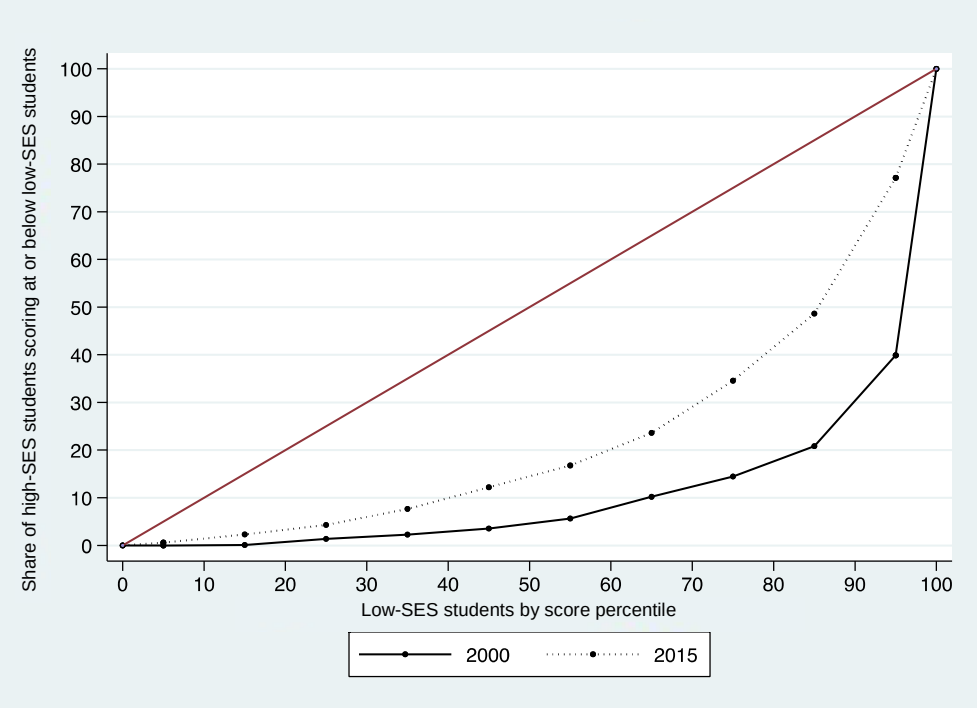
Figure 10. SES-Achievement Gap Calculated with Alternative Point Estimation Method, Birth Cohorts 1954-2001



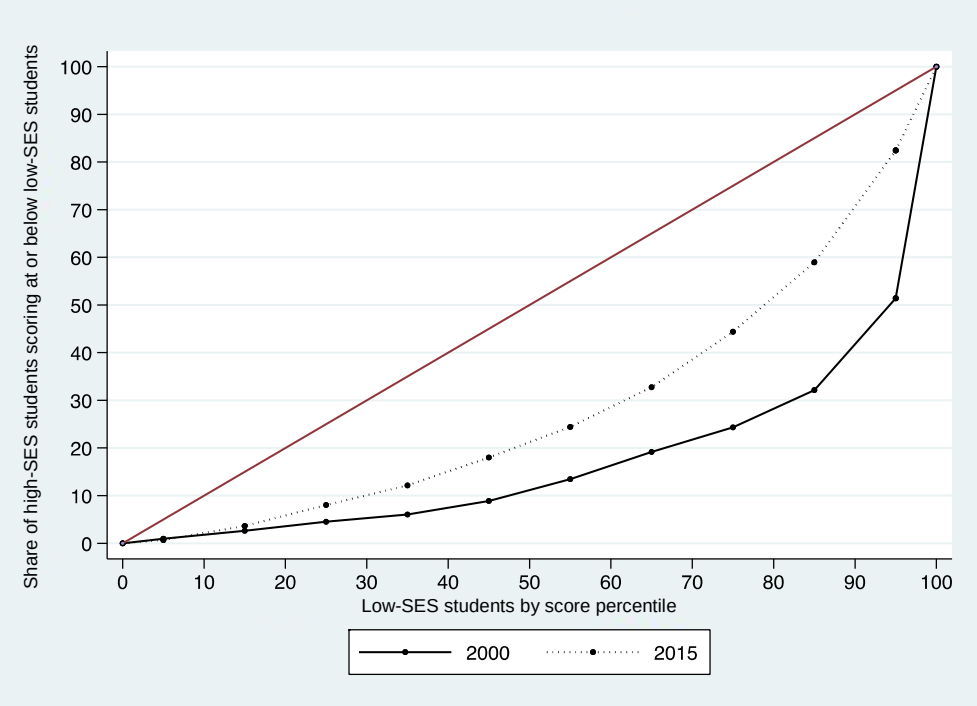
Notes: Difference in achievement estimated at the 75th and the 25th percentile of the SES distribution based on cubic estimation function of the SES-achievement relationships. See Figure 2 for data and additional information.

Figure 11. Ordinal Analysis of Change in the SES-Achievement Distributions, PISA 2000 and 2015

Panel A: Bottom quarter vs. top quarter of the SES distribution



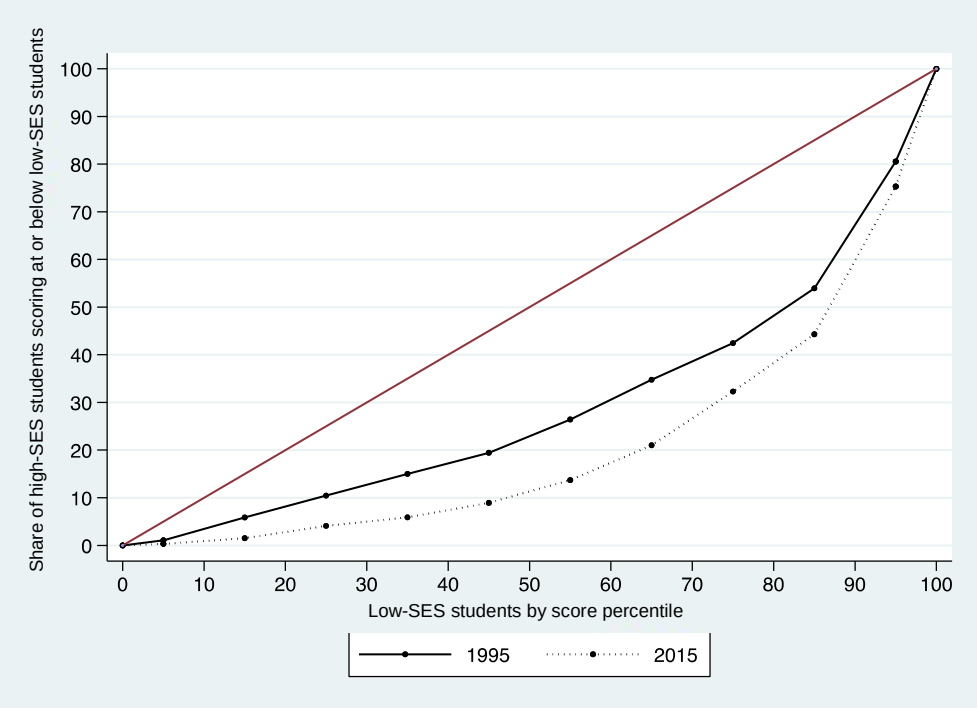
Panel B: Bottom quarter vs. top half of the SES distribution



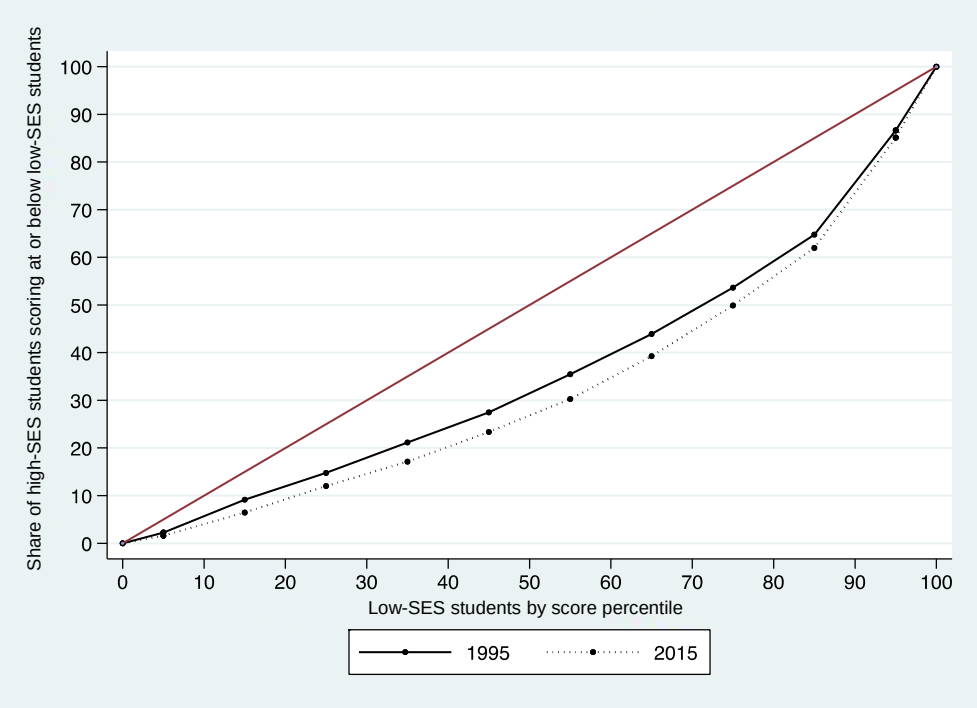
Notes: Horizontal axis: students in the low-SES group lined up along their math achievement distribution. Vertical axis: share of students in the high-SES group who score at or below the respective math achievement of the low-SES percentile.

Figure 12. Ordinal Analysis of Change in the SES-Achievement Distributions, TIMSS 1995 and 2015

Panel A: Bottom quarter vs. top quarter of the SES distribution



Panel B: Bottom quarter vs. top half of the SES distribution



Notes: Horizontal axis: students in the low-SES group lined up along their math achievement distribution. Vertical axis: share of students in the high-SES group who score at or below the respective math achievement of the low-SES percentile.