LONG-RUN TRENDS IN THE U.S. SES—ACHIEVEMENT GAP

Eric A. Hanushek
(corresponding author)
Hoover Institution
Stanford University
Stanford, CA 94305
hanushek@stanford.edu

Jacob D. Light
Department of Economics
Stanford University
Stanford, CA 94305
jdlight@stanford.edu

Paul E. Peterson
Department of Government
Harvard University
Cambridge, MA 02138
ppeterso@gov.harvard.edu

Laura M. Talpey
Hoover Institution
Stanford University
Stanford, CA 94305
lauratalpey@gmail.com

Ludger Woessmann
Ifo Institute
University of Munich
Munich, Germany
woessmann@ifo.de

Abstract

Rising inequality in the United States has raised concerns about potentially widening gaps in educational achievement by socioeconomic status (SES). Using assessments from LTT-NAEP, Main-NAEP, TIMSS, and PISA that are psychometrically linked over time, we trace trends in SES gaps in achievement for U.S. student cohorts born between 1961 and 2001. Gaps in math, reading, and science achievement between the top and bottom quartiles of the SES distribution have closed by 0.05 standard deviation per decade over this period. The findings are consistent across alternative measures of SES and subsets of available tests and hold in more recent periods. At the current pace of closure, the achievement gap would not be eliminated until the second half of the 22nd century.

https://doi.org/10.1162/edfp_a_00383

© 2022 Association for Education Finance and Policy

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, both states and the federal governments enacted a wide range of new education programs designed to enhance the human capital of children born into poor and otherwise disadvantaged households. Surprisingly little, however, is known about the educational outcomes of disadvantaged students over the subsequent decades. In this paper, we provide evidence on trends in achievement gaps between children raised within families of high and low socioeconomic status (SES) for cohorts born from 1961 to 2001. Our main finding is that the SES–achievement gap has fallen modestly over these four decades, suggesting potential for some improvement in intergenerational mobility but an improvement that will evolve slowly over the remainder of the century.

For good economic reasons, 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). In advanced industrial societies, cognitive skills as measured by student performance on standardized tests 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), implying that the U.S. labor market also punishes the lack of cognitive skills more than other developed countries.

The role of families in contributing to achievement differentials is indisputable. Since the Coleman Report (Coleman et al. 1966), families have been seen as a crucial, if not dominant, input to children's development of cognitive skills. Our interest focuses on the intertemporal dynamics of this relationship. Because the family–achievement linkage impedes intergenerational mobility, it is useful to investigate the extent to which the policies initiated in the War on Poverty and related programs have been successful at reducing SES–achievement gaps over time. Although this analysis cannot pinpoint the causes of any changes, it shows the trends that have occurred.[2]

Tracing the pattern of achievement gaps requires consistent measures of achievement. Because of the idiosyncrasies of different testing regimes, simply linking the results from different tests might show variations in the pattern of achievement unrelated to any fundamental changes in student cognitive skills.

Fortunately, there are four high-quality testing regimes that provide both intertemporal achievement data for nationally representative samples of U.S. students and information about standard proxies for family inputs. We draw upon data from these well-documented surveys that have used established psychometric methods to link achievement in math, reading, and science over time. These assessments are the Long-Term Trend of 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

---

[1] See http://www.lbjlibrary.net/collections/selected-speeches/november-1963-1964/01-08-1964.html.

[2] Surprisingly little research provides well-identified causal estimates of how aspects of the family contribute to child outcomes and to intergenerational mobility. For a review of prior work and causal estimates of the transmission of cognitive skills, see Hanushek et al. (2021).

administered to representative samples of U.S. adolescent students who were born between 1961 and 2001. Tests within each assessment regime were designed to provide achievement measures that can be reliably compared over time.

Tracing the pattern of achievement gaps and of the role of family differences proves to be surprisingly difficult. There is no unified data source that clearly describes the distribution of family inputs and their relationship to the distribution of student achievement, in part because there has been relatively little concern about the measurement of family inputs. Historically, variations in family inputs have been proxied by whatever simple measures of SES are available. Virtually no analysis has documented the causal structure of in-family learning or identified the precise causal impact of each of the multitude of family inputs into the child's educational progress, though careful studies have estimated effects of specific factors in particular places and at particular times.

The four assessment regimes provide information about the family background of each student, albeit not as consistently measured as achievement. That information permits construction of a picture of the distribution of family inputs relevant to each test administration and thus to an estimation of changes in SES–achievement gaps over time.

Using individual data for over one and a half million students, we construct an index of SES based on information about parental education and home possessions of the students for 93 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 for the 77 of these 93 testing occasions for which sufficiently detailed family data are available. We estimate a quadratic trend of the aggregate pattern in SES–achievement gaps over time, controlling for assessment regime, subject, and schooling-level fixed effects.

We find a steady, albeit modest, reduction in the SES–achievement relationship over the past four decades. For our earliest cohort, the SES–achievement gap between the top and bottom SES quartiles (75–25 SES gap) is roughly 0.9 standard deviation (SD). This gap is equivalent to a difference of roughly three years of learning between the average student in the top and bottom quartiles of the SES distribution.[3] Over the four decades we study, we find in our preferred model that the gap narrows at 0.05 SD per decade, closing about a fifth of the initial SES–achievement gap but indicating that at this rate of closure it would take another century and a half to completely close the gap.

In sensitivity and robustness analyses, we show that our results hold up after consideration of a range of methodological issues. A falling SES–achievement gap is observed in multiple sensitivity analyses that explore impacts by subject, testing regime, and specific time period. Further robustness analyses include use of alternative approaches for measuring SES, of an alternative point estimation approach to our preferred group calculation approach, and of an alternative analysis that considers the ordinal nature of the underlying achievement data.

---

[3] Differences across grades and ages on the vertically linked NAEP tests support the rough rule of thumb that one standard deviation of achievement is equal to three to four years of schooling; see Hanushek, Peterson, and Woessmann (2012a, 2012b). Note, however, that this correspondence has not been extensively researched and is likely to vary by grade level, position in the test distribution, and other factors.

The next section reviews the prior literature on trends in SES–achievement gaps. Section 3 describes our achievement data. Section 4 discusses the measurement of SES–achievement gaps and section 5 the estimation of trends in these gaps. Section 6 reports our main evidence on trends in student achievement gaps and levels. Section 7 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

Ever since Neff's (1938) pioneering study, broad and extensive research has consistently shown that children from different families reach different achievement levels and that these family differences are highly correlated with family SES. Coleman et al. (1966), in their seminal study reported in *Equality of Educational Opportunity,* found parental education, income, and race to be highly correlated with student achievement in their cross-sectional data while finding that school factors were much less significant. Subsequent research into family factors has consistently confirmed these early findings on the role of families (Smith 1972; 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).[4] 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; Egalite 2016).

The precise definition 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 2012, p. 4). The common interpretation is that the correlations between SES and achievement primarily represent systematic differences in parent–child interactions and in parenting styles. As Cheng and Peterson (2019) discuss, research has pointed to a variety of specific potential mechanisms by which higher SES might operate including such things as introducing a larger vocabulary (Hart and Risley 1995, 2003), superior parenting practices (Hoff 2003; Guryan, Hurst, and Kearney 2008; Doepke, Sorrenti, and Zilibotti 2019), access to a more enriched schooling environment (Altonji and Mansfield 2011), and less exposure to violent crime (Burdick-Will et al. 2011). Many suggest that these and other childhood and adolescent experiences may contribute to SES disparities in academic achievement (Kao and Tienda 1998; Perna 2006; Goyette 2008; Jacob and Linkow 2011). But no one argues that any of these specific factors provide the basis for a comprehensive measure of family inputs.

In empirical analyses, measures of SES are ordinarily based upon data availability rather than conceptual justification.[5] 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, e.g., Gould, Simhon, and Weinberg 2019). Rather, the general approach is to look for more readily available indicators of

---

4. Similar relationships are also found in international studies. For example, an analysis for Britain shows stability of SES impacts over a ninety-five-year period (von Stumm, Cave, and Wakeling 2022).
5. Marks and O'Connell (2021) make the argument that common SES measures, including those used in OECD (2018), are conceptually and empirically weak compared with measures of cognitive ability/genetic inputs, but again this argument comes down to availability of data and differences in analytical focus.

persistent cultural and economic differences across families as proxies for the bundle of educational inputs of families. The standard list includes parental education, occupation, earned income, and various items in the home (National Center for Education Statistics 2012; Sirin 2005). Importantly, these separate measures tend to be highly correlated with each other so that missing data on some of these elements is not overly damaging in characterizing the distribution of SES. At the same time, the correlations make separating their individual impacts on learning difficult and thus make the identification of their relative importance problematic.

Three studies take a similar analytical approach to ours. They develop composite measures of family SES and then relate them to the recent pattern of student achievement measured by one of the available repeated testing regimes. The first study, by the Organization for Economic Co-operation and Development (OECD) (2018), estimates the change in the SES–achievement gap between 2000 and 2015 as traced through the consistent set of psychometrically linked tests in math, science, and reading of the PISA assessments. The OECD measure of SES is its index of Economic, Social, and Cultural Status (ESCS) that aggregates data from students on their parents' education, their parents' occupation, and an inventory of items in their home. The study gauges changes in the SES–achievement connection by identifying changes in the socioeconomic 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 because parental background explains less of the variation in student achievement.

In the second analysis, Broer, Bai, and Fonseca (2019) use the 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). While more home resource measures are available, those used in their study were restricted to these two in order to maintain exact comparability over time and across countries. They compute the distribution of their SES index for each country-year observation using predetermined weights for the underlying elements. Because 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. They find that the SES–achievement gap for the United States declines significantly in science but does not change significantly in math.

In the third analysis, Bai, Straus, and Broer (2021) develop SES trends in math for eighth graders in U.S. states and in the nation as a whole from the Main-NAEP for 2003–17. They construct an SES measure comparable to the measure used in the previous TIMSS analysis, using a constant-weight index of survey items for parent education, books in the home, and individual- and school-level participation in the National School Lunch Program. They conclude that the 75–25 SES gap remained constant for the United States as a whole and for 34 of the 50 states, while the gap widened for 14 states and narrowed for 2 states.

A more ambitious study by Chmielewski (2019) traces achievement gaps as estimated by several family input indicators with data from an array of international surveys. The study combines data from one hundred countries on international tests conducted between 1964 and 2015 to estimate SES–achievement gaps by comparing students estimated to be at the 90th percentile to those estimated to be at the 10th percentile. It relies chiefly upon parental education as the SES indicator but also separately analyzes relationships between achievement and parental occupation and books in the home. It 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 the use of tests that are not psychometrically linked and reliance on broadly defined parental-education categories that require extensive extrapolation outside observed SES levels.[6]

In a recent analysis of psychometrically linked tests, Shakeel and Peterson (2022) report heterogeneity in achievement gains of U.S. students over the past half century by both ethnicity and by SES. Their SES index relies upon indicators of parental education and possessions in the household. They find slightly greater average achievement progress for those in the bottom than in the top quartile of the SES distribution.

Two studies estimate the SES–achievement trend by use of current family-income indicators, although both suffer from weak or inconsistent measures of income that are used as the single dimension of SES. Reardon (2011) estimates income-achievement relationships for sampled students in twelve separate surveys with very different student testing and uses the varying birth years for students across surveys to provide information about the dynamics of SES–achievement gaps. He concludes that there has been a significant increase in SES–achievement gaps over the past quarter century, but that finding appears likely to result from measurement errors both in income and in achievement.[7] Hashim et al. (2020) estimate the SES–achievement gap using NAEP data from 1990–2015 and conclude that gaps have declined. The analysis infers individual income–achievement relationships from income indicators for geographic areas.

In somewhat related work, numerous studies look at the black–white test score gap in the United States (see, e.g., Grissmer, Kirby, Berends, and Williamson 1994; Grissmer, Flanagan, and Williamson 1998; Jencks and Phillips 1998; Hanushek 2001; Magnuson and Waldfogel 2008; Reardon 2011). Because SES backgrounds of black and white students differ markedly, changes in the black–white test score gap provide a partial window on trends in the SES–achievement gap. But the correlation between race and SES has been declining (Wilson 1987, 2011, 2012) and black students constitute only about 16 percent of the school-age population (Rivkin 2016). Thus, the pattern of black–white achievement gaps only provides a limited picture of changes in the overall SES–achievement gap for the United States.

In sum, scholars have used a wide variety of surveys and a range of SES indicators to explore the trend in the size of the SES–achievement gap. One study relies

---

[6] The difficulties of extrapolation of achievement data when there are a limited number of categories in the SES measure are discussed in section 7 (below). They also enter implicitly in our sample selection criteria described in section 4 (below). A recent approach to more reliable linking of the different international tests is found in Majoros, Rosén, Johansson, and Gustafsson (2021).

[7] For a discussion of the measurement issues, see Hanushek et al. (2022).

on PISA (OECD 2018), another on TIMSS (Broer, Bai, and Fonseca 2019), two have used the Main-NAEP (Hashim et al. 2020; Bai, Straus, and Broer 2021), one analyzes multiple tests (Shakeel and Peterson 2022), and two have used a broad range of non-psychometrically linked tests (Reardon 2011; Chmielewski 2019). Most have estimated SES with an index based upon information available from the survey, but two rely solely on income indicators (Reardon 2011; Hashim et al. 2020). Estimates of trends range from downward trending gaps to no significant change to steep increases.

Our work builds on these prior works by estimating trends in the SES–achievement linkage using all available information from psychometrically linked tests that are designed to track temporal change. We expand the analysis to include the full set of the four relevant test regimes designed for trend analysis of cognitive skills of U.S. students. Within this context, we investigate the time pattern of SES–achievement gaps over four decades and assess the potential impact of alternative ways of measuring SES.

3. LONGITUDINAL ACHIEVEMENT DATA

The four longitudinally designed testing regimes used in our analysis provide test performance for representative samples of U.S. adolescents over multiple years. The tested subjects, which vary by assessment regime, include mathematics, reading, and science.[8] Each test is designed to be comparable over time by utilizing psychometric linking based on repeated test items across test waves. All are low-stakes tests: No consequences to any person or entity are attached to student performance, 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 to construct an SES index. Each dataset provides micro data at the student level, making it possible to compare student performance across family SES levels.

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.[9] LTT-NAEP data are available in select years for reading from 1971–2012 and for math from 1973–2012, although the limited information on family background (described below) means that we will be unable to use the earliest years in our analysis of achievement gaps. 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

---

[8] The assessments also vary in a variety of operational details such as whether a sampled student is tested in just one subject (NAEP) or multiple subjects (PISA and TIMSS) at each administration.

[9] 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. Further, 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).

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 and were analyzed under a restricted-use data license.

This highly regarded survey provides the longest available performance history. We use test results obtained at age 13 and age 17, when students are close to leaving secondary school. In our analysis, we use data beginning with the 1961 birth cohort when the survey contains adequate information for estimating SES background. As of that cohort, the survey obtains information on parental education, but data on items in the home used for constructing an SES index are limited, especially in the earlier years.

The age 17 tested population is potentially subject to some varying selection over time with changes in high school graduation rates. High school graduation rates were roughly constant from 1970 to 2000 (Heckman and Lafontaine 2010) and increased by roughly 5 percentage points from 2000 to 2010 (Murnane 2013). In the analysis, we consider the potential impact of this subset of students on the results.

Main National Assessment of Educational Progress (Main-NAEP)

Main-NAEP administers tests of math and reading aligned to the curriculum in grade 8.[10] Begun in 1990 with new administrations of the survey every two to four years, it is designed to provide trend results for representative samples of students in the United States as a whole and for each participating state.[11] Main-NAEP maintains a reputation for reliability and validity similar to LTT-NAEP, but the testing framework is periodically adjusted to reflect changes in school curricula. For each administration of the test, the Main-NAEP sample is approximately 150,000 observations, the large sample necessary in order to have representative samples for each state. Similar to LTT-NAEP, the surveys of background information in Main-NAEP are somewhat limited, particularly in the early years.

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.[12] The tests are 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 to the next. We use the TIMSS eighth-grade math and science tests beginning with the 1981 birth cohort. TIMSS data are available every four years from 1995 to 2015. The U.S. sample includes approximately 10,000 observations for each administration of the test.[13]

---

[10] We exclude Main-NAEP science because eighth-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. We exclude tests administered to students in fourth grade for reasons discussed in footnote 9.

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

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

Although TIMSS is focused on international comparisons, it has ample samples of U.S. students and uses tests that are highly regarded for their psychometric properties. TIMSS has varying detailed background information over time, with family information for recent years being particularly rich.

Programme for International Student Assessment (PISA)

PISA, administered by the OECD, began in 2000. It was originally designed to provide comparisons among OECD countries, but it has since been expanded to a broader set of countries. 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 U.S. sample includes over 5,000 students for each administration of the test.[14] The United States has participated in every wave of the test, allowing us to use national PISA data available every three years from 2000 to 2015.

While the PISA tests are designed to assess the ability of students to apply skills to real-world problems, the overall performance on these tests is highly correlated with the curriculum based testing in TIMSS (Loveless 2017). Results are not available for reading for the 1991 birth cohort because of test administration problems. The family background surveys are consistently highly detailed.

Summary of Test Information

Table 1 provides a schematic of the assessment data that are available across years in the four surveys. The coverage in the earlier years is clearly thinner than in the later years because only LTT-NAEP is administered before 1990.

In the subsequent analysis, we compile an aggregate distribution of achievement from student-level micro data available for each subject, testing age, and birth cohort for a forty-year period. To equate results across tests, we express achievement in standard deviations for each testing regime, subject, and testing age, normalizing mean achievement in 2000 (or the closest test year) to zero.[15] 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 that can be used in our analyses was administered by LTT-NAEP in reading to a cohort of students born in 1961; the last test in our analysis was administered to students born in 2001. Across this four-decade span, achievement data are available for 1,695,574 students from 44 tests in math, 37 in reading, and 12 in science. Table 2 summarizes for each survey of the number of assessments, subject matter, age or grade level at which students are tested, birth cohorts surveyed, and number of observations. We use 77 of the potentially available 93 separate test-subject-age/grade-year observations,

---

[13] 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. We exclude tests administered to students in fourth grade for reasons given in note 9.

[14] The PISA testing of sampled students differs from that for Main-NAEP. In PISA, sampled students take each subject assessment; in NAEP, sampled students take just one subject assessment.

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

Table 1. Surveys and Subjects by Test Date, 1971–2015

	LTT-NAEP				Main-NAEP		PISA			TIMSS	
	13-year-olds		17-year-olds		Eighth graders		15-year-olds			Eighth graders	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Science	Math	Science
1971		(X)		(X)							
1972											
1973	(X)		(X)								
1974											
1975		X		(X)							
1976											
1977											
1978	X		X								
1979											
1980		(X)		X							
1981											
1982	X		(X)								
1983											
1984											
1985											
1986	X		(X)								
1987											
1988		X		X							
1989											
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: Test identifiers in parentheses indicate assessments where the SES information is insufficient to calculate 75–25 gaps. LTT-NAEP = Long-Term Trend of the National Assessment of Educational Progress; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

Table 2. Summary of Assessment Data

				Observations by Test and Subject			
Test	**Age/Grade**	**Birth Cohorts**		**Math**	**Reading**	**Science**	**Total**
LTT-NAEP	age 13	1961–1999	Waves:	11	11		22
			Students:	78,210	71,430		149,640
LTT-NAEP	age 17	1961–1995	Waves:	11	10		21
			Students:	77,610	57,870		135,480
Main-NAEP	grade 8	1977–2002	Waves:	10	11		21
			Students:	583,012	664,556		1,247,568
TIMSS	grade 8	1982–2002	Waves:	6		6	12
			Students:	44,074		44,074	88,148
PISA	age 15	1985–2000	Waves:	6	5	6	17
			Students:	26,173	22,390	26,175	74,738
Total			Waves:	44	37	12	93
			Students:	809,079	816,246	70,249	1,695,574

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 mean achievement levels and not for the distribution of individual scores. Sample sizes for the restricted-use NAEP data are rounded to the nearest 10. LTT-NAEP = Long-Term Trend of the National Assessment of Educational Progress; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

with the restrictions in the sample reflecting insufficient family background information for our trend estimation (see below).[16]

Each assessment regime is highly regarded not only for the psychometric properties of its test but also for the care in sampling students and the enforcement of testing protocols. As a result, we have no a priori reason to believe any one test is more reliable or valid than any other, and we use all possible information about trends in achievement gaps for the measured cognitive skills.

4. MEASURING SES–ACHIEVEMENT GAPS

Estimating achievement differences across the SES distribution requires a measure of family SES that adequately depicts the underlying distribution of the population and that can be applied at the individual student level. We construct an SES index based on student-reported information of parental education and home possessions, information that is provided within all four assessment surveys included in this analysis.

The measurement task is complicated by the varying nature of data availability associated with the different underlying assessments. The survey questions typically ask about parental educational attainment in categories and home resource questions as binary responses to presence of items or, in some cases, categories, such as different ranges of number of books in the house (0–10, 11–25, etc.). Across the various surveys that stretch for four decades, the indicators of family background change both in specifics and in interpretation. The specific items differ across the individual

---

[16] Further, we consider only birth years after 1960. Although there were two earlier math and reading assessments in the LTT-NAEP, the survey background information would not support the estimation of SES gaps. Data for the 77 observations of the 75–25 gap that are used in the main analysis can be found in appendix table A3, available in a separate online appendix that can be accessed on Education Finance and Policy's Web site at https://doi.org/10.1162/edfp_a_00383.

Table 3. Sample of Survey Questions on Items in the Home and Parent Education, Main-NAEP 1990, and PISA 2000

Home items: 1990 NAEP	Home items: 2000 PISA
Do you have an encyclopedia in your home?	Do you have a dishwasher in your home?
Is a newspaper delivered regularly to your home?	Do you have your own room?
Are magazines delivered regularly to your home?	Do you have educational software in your home?
Do you have more than 25 books 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?
	Do you have books of poetry in your home?
	Do you have works of art in your home?
	How many cell phones do you have in your home?
	How many televisions do you have in your home?
	How many calculators do you have in your home?
	How many computers do you have in your home?
	How many musical instruments do you have in your home?
	How many cars do you have at your home?
	How many bathrooms do you have in your home?
	How many books do you have in your home?
Parent education: 1990 NAEP	**Parent education: 2000 PISA**
Did not finish high school	Did not go to school
Graduated high school	Completed primary education
Some education after high school	Completed lower secondary education
Graduated college	Completed upper secondary (vocational)
	Completed upper secondary (tertiary entry)
	Completed tertiary

Notes: Main-NAEP (National Assessment of Educational Progress) 1990 and PISA (Programme for International Student Assessment) 2015 chosen for expositional purposes only as examples with a low and high number of information categories.

testing regimes, and they are sometimes changed across years within each assessment regime. Importantly for this analysis, the items also differ in the granularity with which parental education is measured and in the scope of items in the home that are included.

The summary of questions for the 1990 Main-NAEP survey and for the 2000 PISA survey identified in table 3 provides an indication of the variability in the underlying survey data. While the 2000 PISA survey inquired about Internet access and home computers, those questions were not asked by the Main-NAEP ten years earlier when such items were rare. Similar differences in scope and categorization of paternal education occur. In some administrations of the surveys, students were asked about their parents' education in detail; in others, the question was phrased more broadly.

Table 4 summarizes the quantitative differences in survey detail among the different test regimes and across time within each regime. This clear variation in detail actually understates the differences across surveys since, for example, the number of books in each category can change over time.

Two general conclusions about measuring SES emerge from this overview of the survey data. First, there is no set of common items that are measured across time and

Table 4. Number of Categories and Items of Student Surveys by Earliest and Latest Survey for Each Assessment Regime

	Parental Education Categories		Books in the Home Categories		Number of Home Items	
	Earliest	Latest	Earliest	Latest	Earliest	Latest
LTT-NAEP-age 17	4	4	2[a]	4	3	4
LTT-NAEP-age 13	4	4	2[a]	4	3	4
Main-NAEP	4	4	2[b]	4	3	6
TIMSS	5	5	5	5	16	8
PISA	6	6	7	7	18	24

Notes: LTT-NAEP = Long-Term Trend of the National Assessment of Educational Progress; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

[a] Book categories of less than 25 books or greater than 25 books.

[b] Book categories of have books or do not have books.

testing regimes.[17] Second, even if there were a common set of measures, it would not necessarily be optimal to create a fixed SES index based on them, because the meaning of these indicators changes over time. Having a magazine delivered to the home provides a different picture of the family in 1975, in 1990, and in 2015. Similarly, changing educational attainment since the 1960s has changed the socioeconomic implications of having, for example, a high school diploma as the highest level of education completed.

Our preferred SES index is constructed as the first principal component of a full vector of dummy variables representing all available home resource questions and a vector of dummies corresponding to each level of education for the parents' level of education.[18] The principal component analysis, similar to that used by OECD (2018), reduces the dimensionality of the data while preserving important variation across the underlying set of factors.[19] We estimate a separate SES index for each testing regime and for each year within each testing regime. For LTT-NAEP, we estimate separate SES indices for 13-year-old and 17-year-old students.

We develop separate indices for each testing regime because survey questions significantly differ across testing regimes. Similarly, we develop separate indices within testing regimes across years to control for changes in the link between certain home resources and SES. The surveys span decades over which technology evolved substantially. Technologies that were niche or luxury goods in the 1990s, such as home Internet or owning a cell phone, became commonplace in recent years. Other technologies,

---

[17] Even within testing regime, there are limited common survey items. For example, Broer, Bai, and Fonseca (2019) wished to create a fixed-weight index of common items that covered the twenty-year span of TIMSS surveys and ended up with an index using just four measures: highest parental education, presence of a computer, availability of a study desk, and books in the home. Even within these limited categories, the survey items for computer presence changed over time.

[18] If the student provides levels of education for more than one parent, we use the highest level of education among the student's parents. We drop observations with missing answers for questions, although the incidence of missing items is not severe. With a few exceptions, full data are available for 95 percent or more of students.

[19] The underlying data for our SES index in our preferred estimation are categorical. An alternative to standard principal components analysis (PCA) when there is a concern about underlying normality assumptions is multiple correspondence analysis (MCA). When we estimate the main model with MCA, the pattern is not qualitatively different from using PCA (using STATA *mca* and *pca* commands). We rely on the PCA estimates because they facilitate comparisons of alternative ways to construct the SES index as described subsequently in the Alternative Construction of SES Measures section.

such as record or CD players, have transitioned from popular to obsolete. Education levels have also dramatically risen since the 1960s.[20] Estimating separate SES indices ensures that our estimates are not biased by the changing importance of certain home resources across years.

This estimation provides the fewest a priori restrictions on the education and home item data, but it is not the only possible approach. Instead of using the general principal components approach, it would, for example, be possible to create a predetermined index of items directly (Broer, Bai, and Fonseca 2019), to convert the education or books in the home categories into numerical values, or to provide different weights for the major components. In the robustness analysis of the Alternative Construction of SES Measures section, we consider the potential impact of a range of alternative constructs for SES, but these alternatives do not alter the overall results for estimated trends in SES gaps.

Our main analysis focuses on the SES–achievement gap measured by the difference in average achievement between students in the top and bottom quartiles of the distribution of the SES index. That is, we compare the 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 SES percentile. For expositional purposes, we refer to this as the 75–25 SES gap.

Survey questionnaires in general collect information about subjects that more precisely discriminates among individuals near the middle of the distribution of the population than those at the extremes. As a result, those at the extremes of a distribution, especially those in the right-hand tail, are often bundled together into broad categories that include a large percentage of all observations, making it difficult to estimate reliably differences within the category. For example, the category "college degree or more" might include close to half the sample. When a survey has these kinds of broad categories and only a few questions (see tables 3 and 4), an SES index may provide an imperfect estimate of differences in family background for those in the right-hand tail of the distribution. Potentially, the left-hand tail could pose similar problems, but in practice the questionnaires generate information that allowed for fairly precise delineation of variation within that tail.

These aspects of the data can lead to a lack of information about achievement in the right-hand tail of the SES distribution. The coarseness of the survey questions, particularly in the earlier LTT-NAEP and Main-NAEP assessments, means that the SES values estimated by the principal component are themselves categorical and do not distribute themselves smoothly over the entire range of the distribution. In some assessment years, the largest value of the observed SES values may include a broad range of individuals and may not give sufficient information to distinguish scores of those in the top quartile of the SES distribution from those at lower points in the distribution.

The set of survey items in table 3 illustrates the issue. In 1990, the first year for Main-NAEP, students were asked to place their parents within one of four education categories and to respond about whether they had each of only four items in their home.

---

[20] The population aged twenty-five and older completing high school rose from 40.1 percent in 1960 to 90.9 percent in 2020. Completing a bachelor's degree went from 7.7 percent to 37.5 percent over the same period (U.S. Department of Education 2020).

As a consequence, when we construct our SES index based on these two measures, over a quarter of the students that Main-NAEP tested in 1990 are identified as being in the top SES category, making it impossible to obtain a precise estimate of achievement for the top quartile of the distribution. Similar problems emerge for other surveys in early administrations of both the Main-NAEP and LTT-NAEP.

Accordingly, when 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. The exclusion rule reduces our observations to 77 out of the potential 93 observations, but it allows us to be more confident about identifying achievement patterns in the tails of the distribution. In other words, because we are making inferences about relative achievement in the tails of the SES distribution, we do not want to begin by imposing fixed distributional assumptions on the achievement in the tails of the SES distribution. Table 1 identifies the specific assessments that are excluded from the analysis of 75–25 SES gaps. We are, however, able to assess the potential bias in the trend estimation from this sample selection rule. 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 89 of the 93 observations; this does not affect our overall results.

The categorical nature of the derived 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 follow convention by using local linearization to interpolate average achievement between that observed for the closest SES category immediately above and closest immediately below the desired cut point.[21]

5. ESTIMATING TRENDS IN ACHIEVEMENT GAPS

The four separate assessment regimes—LTT-NAEP, Main-NAEP, TIMSS, and PISA—are internally consistent over time, but they vary from each other in a variety of details, including relationship to the curriculum, testing philosophy, and sampling frames. We assume that each regime provides a valid and consistent measure of knowledge in each tested domain despite these variations. Differences in estimated trends among them may also be a function of normal sampling error.

Because the assessments include students of differing ages in various years, we put the data on a common basis for trend analysis by comparing students based on their birth year. To identify the aggregate trend in achievement gaps across birth cohorts, the estimation combines results from all assessments but includes indicators for assessment regime, subject, and (in LTT-NAEP) age.[22] The fixed effects for the four testing regimes take out any time-invariant impact of regime-specific, schooling level-specific, and subject-specific characteristics on the trend-line estimation. Thus, the trend-line estimates rely just on the variation over time within each testing regime and not on the variation among testing regimes.[23]

---

[21] For example, we might observe the average achievement for everybody at the 78th percentile or above of the SES distribution and for everybody at the 70th percentile and above. We assume that scores change linearly within that range and estimate the average achievement of those above the 75th percentile by linear interpolation between the two known average achievement levels.

[22] For schooling level, we distinguish between the 17-year-olds in the LTT-NAEP data and the younger ages (13–15 years) found in other samples. For exposition, we frequently refer to the different schooling levels simply by age.

We first calculate the mean performance at the top and bottom of the SES distribution: $\overline{O}_{isah}^{t}$ is average achievement in standard deviations for each survey i by subject s, testing age/schooling level a, and birth cohort t for the high SES group of students with SES > percentile h; and $\overline{O}_{isal}^{t}$ is average achievement for the low SES group of students with SES < percentile l. The relevant achievement gap is then:

$$\Delta_{isa}^{t} = \overline{O}_{isah}^{t} - \overline{O}_{isal}^{t}. \tag{1}$$

We estimate the performance trend with a quadratic function of birth year:

$$\Delta_{isa}^{t} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \delta_i + \gamma_s + \lambda_a + \varepsilon_{isat}, \tag{2}$$

where δ_i, γ_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 SES–achievement gaps.

In our main analysis, we focus on the SES–achievement gap as depicted by the achievement difference between students in the top and bottom quartiles of the SES index distribution ($h = 75$ and $l = 25$), but we also report trends in other measures of disparities in our specification analyses. 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.

6. TRENDS IN ACHIEVEMENT GAPS

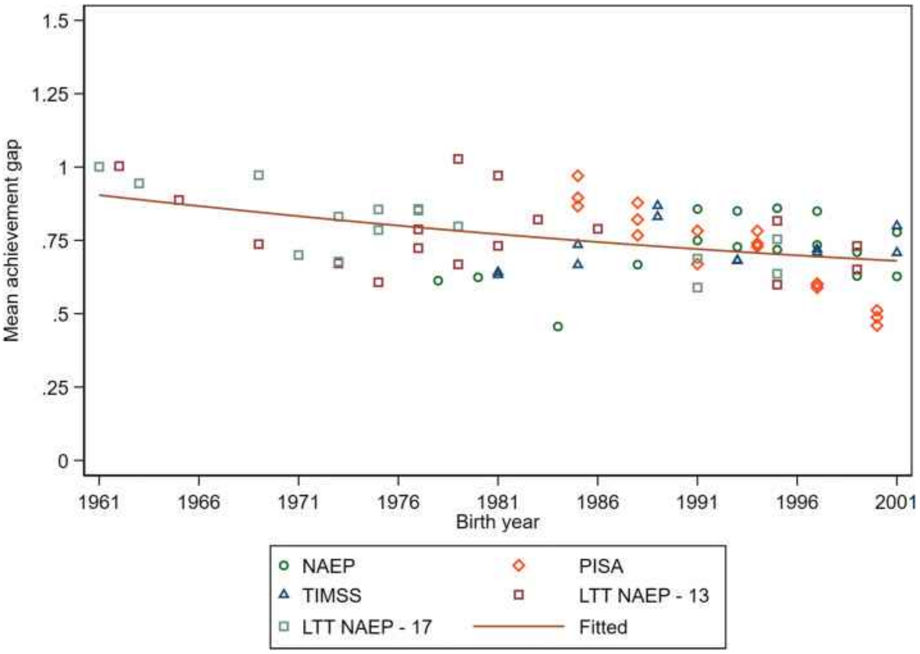
Our preferred estimate shows a modest but steady decline in the SES–achievement gap for the birth cohorts 1961–2001. Figure 1 plots the underlying data points for each assessment regime along with the quadratic trend line. The trend line is not significantly different from linear and shows a linear decline of 0.053 SD per decade. This trend line is based on within-regime data and does not use any between-test regime variation. In the trend estimation, the linear and quadratic terms of the birth year (α_1 and α_2 in equation 2) are each insignificantly different from zero but jointly highly significant [$F(2, 69) = 5.39$, $p < 0.01$]. (Note that details of the estimated models and statistical tests are summarized below in table 5. Annual changes in SD are reported in tables, but for ease of presentation the text discusses changes in SD by decade.)

Trends are quite similar for math and reading performances (figure A1, available in the online appendix).[24] The 75–25 gap for math shows a slightly larger decline over time (see table 5 for details), although the trend differences between reading and math are not significantly different in the simple linear specification. The aggregate trend shown in figure 1 does not mask significantly different trends across subjects.

---

[23] We also estimate models with regime-by-subject fixed effects, but these fully saturated models do not generate results significantly different from our preferred model.

[24] For clarity, we leave out the individual data points in all depictions except figure 1. Markers on the trend lines indicate birth cohorts for which there are data. Because of the limited number of science test observations (see table 2), we do not report separate analyses for science, but the science observations contribute to the aggregate trend analysis.



Notes: Achievement difference between the students in the top and bottom quartiles of the SES distribution (75–25 SES–achievement gap). Data include all tests administered by LTT-NAEP (Long-Term Trend of the National Assessment of Educational Progress), Main-NAEP, PISA (Programme for International Student Assessment), and TIMSS (Trends in International Mathematics and Science Study) with sufficiently detailed SES data (see text). 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.

Figure 1. Achievement Gaps Between Top and Bottom Quartiles of the Socioeconomic Status (SES) Distribution

Inspection of the scatter of data points in figure 1 shows some variation in trends across testing regimes. Figure 2 splits the data for each test regime and estimates a test-specific quadratic trend. Consistent with the analysis in OECD (2018), PISA gaps show a sizeable closing of SES gaps for cohorts born between 1985 and 2000. Over the period, gaps fall by 0.32 SD per decade in math, 0.29 SD per decade in reading, and 0.24 SD per decade in science, each representing larger declines in magnitude than we estimate the other testing regimes.

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 for all of the testing regimes combined.[25] We can nonetheless ascertain the sensitivity of our finding to the inclusion of any particular testing regime. We do so by sequentially excluding each of the test regimes and estimating the quadratic

---

[25] The differences in achievement trends for TIMSS and PISA have been noted previously (Loveless 2017). Loveless shows that at the country level the scores are cross-sectionally highly correlated but the patterns over time, while still highly significant, are less correlated. He notes the differences in sampling by age and the differing basis of questions (curricular versus applied), but offers no explanation for the resulting patterns over time. One possibility (suggested by a referee) is changing patterns of grade progression in the age-based PISA test. While difficult to assess completely, we can calculate the trend in achievement gaps for just the tenth and eleventh graders in PISA and find that a linear trend for these is virtually identical to a linear trend in the complete PISA data that included ninth graders.

Table 5. Estimated Trends in Socioeconomic Status (SES)–Achievement Gaps

Assessment Comparisons	n	α_1	α_2	H_0: $\alpha_1 = \alpha_2 = 0$ (p-value)	Linear (α_1) (with $\alpha_2 = 0$)
Preferred 75-25 SES gaps	77	-0.0087	0.0001	0.007	-0.0053^{***}
Reading	30	-0.0040	0.0000	0.268	-0.0044
Math	35	-0.0166	0.0002	0.017	-0.0060^{***}
Excluding individual assessments					
Exclude PISA	60	-0.0156^{***}	0.0277^{**}	0.011	-0.0032^{**}
Exclude TIMSS	65	-0.0060	-0.0006	0.004	-0.0062^{***}
Exclude Main-NAEP	61	-0.0057	-0.0035	0.000	-0.0072^{***}
Exclude LTT-NAEP	45	0.0339^{***}	-0.1436^{***}	0.002	-0.0033
Exclude NAEP LTT-17	62	0.0001	-0.0080	0.162	-0.0038^{*}
Exclude NAEP LTT-13	60	-0.0079	0.0042	0.018	-0.0058^{***}
Alternative gap definitions					
75–50 SES–achievement gap	77	-0.0075	0.0001	0.148	-0.0027^{*}
50–25 SES–achievement gap	93	-0.0018	0.0000	0.027	-0.0030^{***}
70–30 SES–achievement gap	89	-0.0022	0.0000	0.043	-0.0034^{**}
90–10 SES–achievement gap	52	0.0018	-0.0001	0.019	-0.0081^{***}
Testing period					
After 1990	68	0.0078	-0.0401	0.020	-0.0044^{**}
After 1995	54	0.0119	-0.0756^{*}	0.012	-0.0067^{**}
Fully saturated					
75–25 SES gaps (saturated)	77	-0.0077	0.0000	0.003	-0.0053^{***}

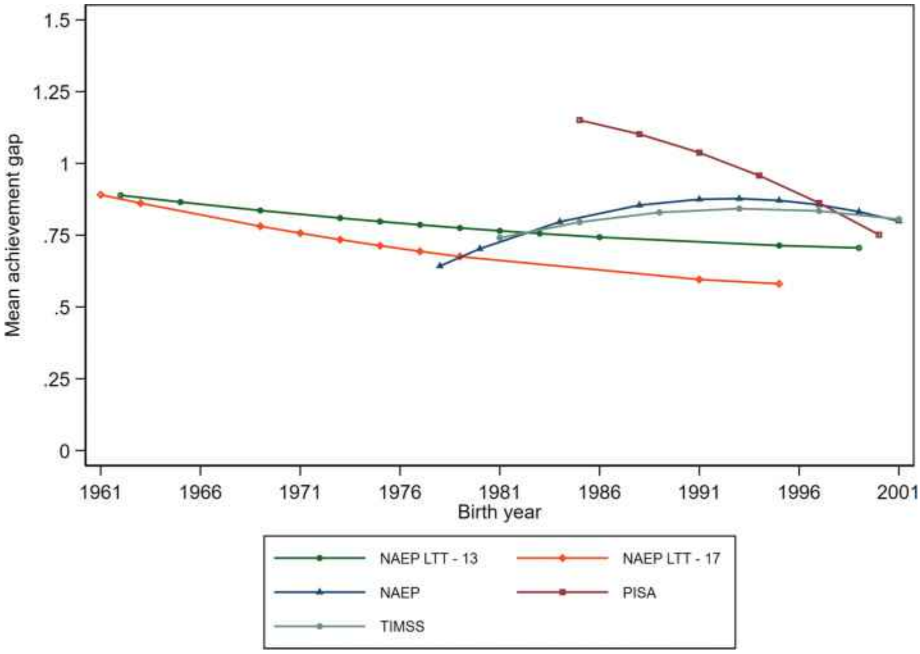
Notes: n = number of observations; α_1 and α_2 are the trend parameters in equation 2; the linear model sets α_2 to zero. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Complete estimates are found in table A1, available in the online appendix. The fully saturated model includes assessment-by-subject-by-age fixed effects. LTT-NAEP = Long-Term Trend of the National Assessment of Educational Progress; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

trend for the remaining three. When PISA observations are excluded from the aggregate trend (leaving sixty observations), the trend parameters are insignificantly different from zero, although the average 75–25 achievement gap still falls by 0.02 SD per decade over the period (table 5 and online appendix figure A2). When we exclude each of the other three test regimes one at a time and reestimate the trend, we also find declining gaps in all instances except for the trend line that excludes LTT-NAEP observations. When LTT-NAEP is excluded, there is a significant curvature to the trend line but there is no overall tilt up or down. (Note that LTT-NAEP is the only series starting before the 1984 birth cohort, and this latter trend excluding LTT-NAEP is estimated over a noticeably shorter time period). If we entirely omit the LTT-NAEP seventeen-year-olds from the analysis, the trend line for the SES gap becomes flatter but remains negative.[26]

There remains the possibility that trends differ in other parts of the distribution. Reardon (2011), for example, finds increasing gaps only in the upper half of the SES distribution (measured by current income). We conduct a similar analysis by comparing

---

[26] As noted, the scores for seventeen-year-olds could partially reflect differential selection due to varying rates of high school completion over time (Murnane 2013). If lower dropout rates increase the share of academically weaker students in high school at age 17, and given the established link between SES and student achievement, one might expect lower dropout rates to increase SES gaps over time. Our estimates, however, go in the opposite direction; the SES–achievement gap declines for the LTT-NAEP seventeen-year-old sample. The lower statistical significance when the seventeen-year-old observations are removed seems to be a reflection of the smaller samples.



Notes: Quadratic trends for 75–25 SES–achievement gaps separately estimated for each test regime. Markers on lines indicate years for which data are available. LTT-NAEP = Long-Term Trend of the National Assessment of Educational Progress; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

Figure 2. Achievement Gaps Between Top and Bottom Quartiles of the Socioeconomic Status (SES) Distribution, by Test Regime

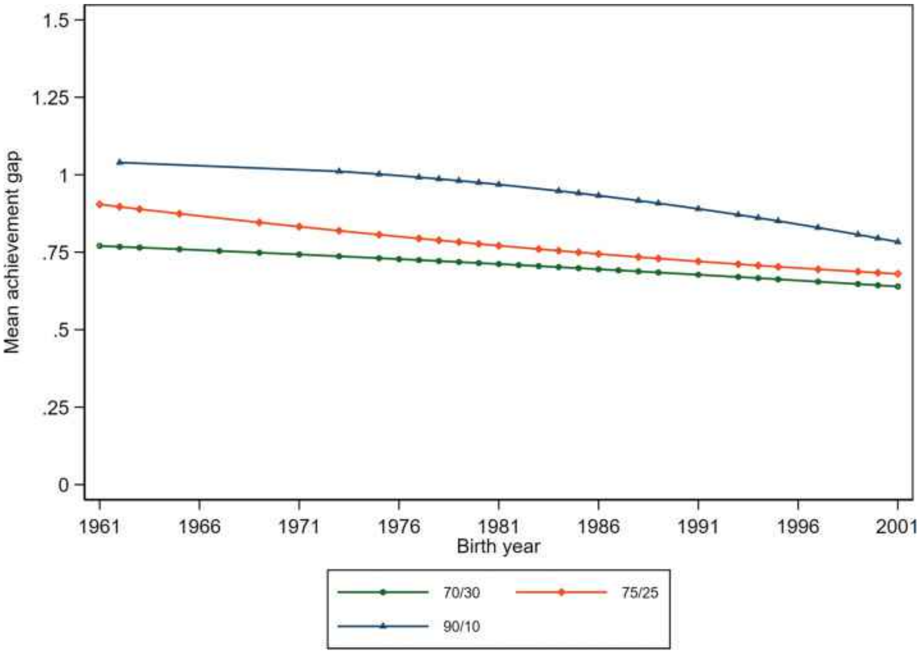
students in the top quartile to students in the bottom half of the SES distribution (75–50 gap) and students in the top half to the bottom quartile of the SES distribution (50–25 gap); see table 5 and online appendix figure A3. Both analyses yield similar estimates of a 0.03 SD per decade decline in SES–achievement gaps.[27]

Our findings of modest declines in the SES–achievement gap are robust to measurement of SES–achievement gaps at other points in the SES distribution. While our primary measure compares achievement at the top and bottom SES quartiles, alternate measures compare achievement at the top and bottom deciles (see, e.g., Corak 2013; Chmielewski and Reardon 2016; Chmielewski 2019). Both to provide a broader set of gap trends and to demonstrate the potential implications of the coarse measurement of SES, we provide estimates of the 90–10 and 70–30 gaps in figure 3.

Estimating the 90–10 SES–achievement gap is only possible for a much smaller subset of our assessments because the available SES data less frequently identify families in the top 10 percent of the distribution. Using the fifty-two test observations with SES information specific to those in the top decile, we estimate a large and statistically significant reduction of 0.08 SD per decade ($p = 0.02$) in the 90–10 SES gap (table 5 and figure 3).

---

[27] The overall dispersion in the distribution of achievement across U.S. students, estimated as a trend similar to equation 2 except that the dependent variable is the unconditional score gap in SD for given percentiles, also narrows somewhat over the four decades. The unconditional interquartile range (the achievement difference between students performing at the 75th and 25th percentiles of the overall achievement distribution) declines by 0.15 SD for birth cohorts 1961–2001.



Notes: Estimated quadratic trends for different SES–achievement gaps. Underlying estimated equations along with number of valid observations for each are found in table 5.

Figure 3. Alternative Gap Calculations: 90–10, 75–25, and 70–30 Socioeconomic Status (SES)–Achievement Gaps

Looking at the 70–30 SES gap is another useful comparison because it increases our observations to 89 of the possible 93 assessments. For this broader sample, there is again a narrowing of the SES–achievement gap, though it is smaller (0.034 SD per decade) than our preferred estimate. The coefficients for the quadratic trend terms are jointly significantly different from zero ($p = 0.04$) within this expanded sample.

Our preferred estimate that relies on just the variation across birth cohorts in performance within test regimes and within subjects can be further refined by including assessment regime-by-subject fixed effects. This fully saturated model yields virtually the same trend estimate as the preferred model (table 5).

A final concern is that the changes we observe may differ across birth cohorts in more complicated ways than captured by our quadratic trend estimation. We have already partially addressed this issue when we exclude the LTT-NAEP data, since the only assessment consistently given before 1990 was the LTT-NAEP. However, when we look at estimated quadratic trends for more recent periods (including the more recent data from the LTT-NAEP), we find significant declines in the SES–achievement gap as measured by tests given 1990–present and 1995–present (see table 5 and the robustness analysis in the Alternative Construction of SES Measures section below).

In summary, the overall picture suggests a steady, if modest, decline in the SES–achievement gap that holds across alternative subsamples of the existing data. Table 5 summarizes the estimated changes in gaps and provides information about the statistical test that the parameters of the quadratic trend are jointly zero. Overall, the alternative specifications provide a consistent picture of the trends. Even looking at subsets of the available testing regimes, there is no indication of the potential increase in achievement gaps that has been identified as a potential result of the widening income distribution (e.g., Duncan and Murnane 2014a).

7. METHODOLOGICAL ALTERNATIVES AND ROBUSTNESS

While the previous section showed the trends in achievement gaps to be highly consistent, it relied upon a common analytical structure. Here we consider alternative structures involving different ways to aggregate the family background data, the use of extrapolated achievement data for separate points in the SES distribution, and the reliance just on the ordinal properties of the achievement scores.

Alternative Construction of SES Measures

Our preferred measure of family SES uses all of the information on parental education and home possessions available in each assessment in binary form (reflecting the available categories for education and books in the home, as well as the binary responses to various home resource items). There are, however, alternative plausible parameterizations of the SES data. The impact of different options and their effects can best be seen using the rich background data available from TIMSS and PISA.[28]

We compare the 75–25 SES–achievement gaps for six alternative ways of constructing the SES index based on survey information about items in a student's home, books in the home, and the highest level of parental education. To show the impact of the various alternatives on the estimated trend in achievement gaps, we begin our trend lines with the 1981 birth cohort, the first assessed in TIMSS. We consider different ways of aggregating the parental education, the books in the home, and the home items data. In all cases, we consider the first principal component as our SES measure, and we produce quadratic trends in achievement gaps across the PISA and TIMSS observations.

The alternatives include transforming the categorical parent education data into a simple linear measure of years of school attainment; transforming the books in the home categorical data into a linear measure of total number of books in the home; collapsing the items in the home into a single category (through a preliminary principal components analysis before combining with the parent education and books in the home data;[29] and estimating a common PCA across time for each test regime.[30]

Specifically, we compare:

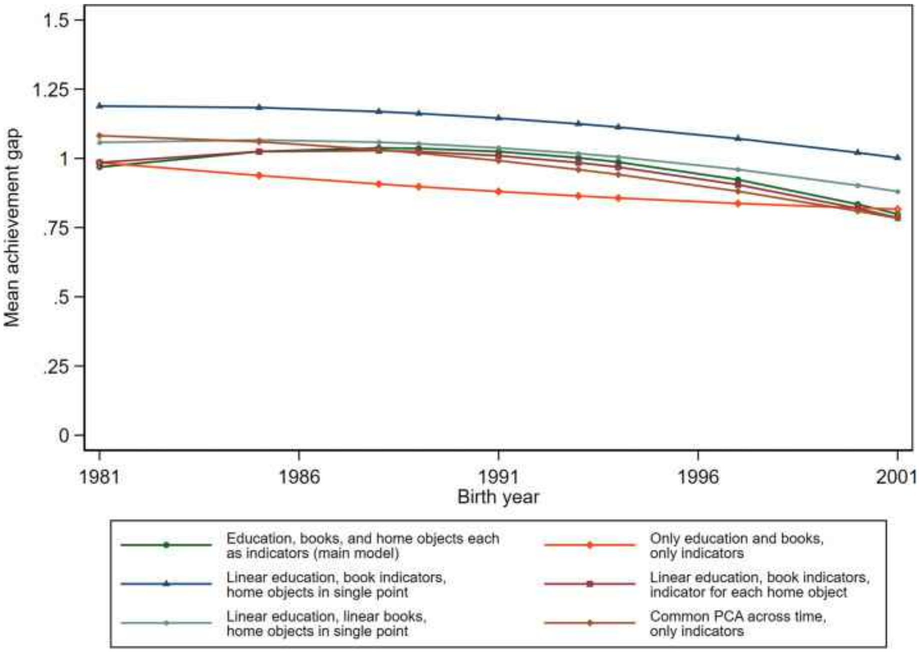
1. Parent education, books in the home, and other items in the home using binary (categorical) measures (our preferred specification);
2. Linear parent education, categorical books in the home, and a single home items index from the first principal component in a preliminary PCA;

---

[28] The paucity of SES indicators in LTT-NAEP and Main-NAEP (see table 4) limits the utility of this robustness check for these surveys.

[29] Earlier analysis (not shown) also divided the various items in the home into learning objectives (such as computers) and wealth objects (such as number of cars). Including these differences into the construction of the SES measure does not, however, lead to different conclusions, and this approach was dropped.

[30] While we prefer separate PCA calculations for each time for the reasons given in section 4, changing survey questions and changing interpretation of them over time could introduce differential measurement error in SES that biases the estimation of trends in achievement gaps. By constraining the SES measure to a common PCA over time for each test regime, we can address these concerns.



Notes: Quadratic trends in the 75–25 SES–achievement gaps estimated for alternative ways of calculating SES. See text for description of the alternatives and table 6 for details of the estimated parameters. TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

Figure 4. Socioeconomic Status (SES)–Achievement Gaps with Alternative Construction of the 75–25 SES Distribution, TIMSS and PISA

3. Linear parent education, linear books in the home, and a single home items index from the first principal component in a preliminary PCA;
4. Categorical parent education and books in the home without including other items in the home;
5. Linear education with categorical books in the home and other items in the home;
6. A common PCA across time (instead of varying with each assessment).[31]

Figure 4 displays the resultant quadratic trend lines for the twenty-nine observations of 75–25 achievement gaps in the PISA and TIMSS data. As figure 2 previously showed, the separate TIMSS trends (slight curvature) and PISA trends (strongly decreasing gaps) differ somewhat over this sample period.[32] When these two test regimes are combined, the prior trend in the main estimates appears.

Although there are small differences in curvature across the six alternatives, the modest decline in achievement gaps is consistent across the alternative constructions of the SES measure (for details on the estimates, see table 6). Thus, we conclude that the pattern in SES trends presented in our preferred results is not driven by any specific approach to construction of the SES measure or by differential measurement error

---

[31] Using the same predetermined index for all observations is the approach of Broer, Bai, and Fonseca (2019) and Bai, Straus, and Broer (2021).

[32] If we include the trends in LTT-NAEP and Main-NAEP (using our preferred measure of SES) in calculating the overall trends for this period, the plots across the six alternatives uniformly show a declining pattern that does not significantly depart from linear.

Table 6. Alternative Estimates of Trends in Socioeconomic Status (SES)–Achievement Gaps: Varying Principal Components Analysis (PCA) Inputs and Point Estimates of Gaps

SES Calculations	n	α_1	α_2	H_0: $\alpha_1 = \alpha_2 = 0$ (p-value)	Linear (α_1) (with $\alpha_2 = 0$)
A. Alternative PCA inputs					
Preferred SES measure	29	0.1015^{*}	-0.0014^{**}	0.003	-0.0103^{***}
Linear education, linear books, home objects in single point	29	0.0449	-0.0007^{*}	0.000	-0.0098^{***}
Linear education, book dummies, home object dummies	29	0.0870^{*}	-0.0012^{*}	0.002	-0.0116^{***}
Linear education, book dummies, home objects in single point	29	0.0296	-0.0005	0.000	-0.0100^{***}
Just books and education dummies	29	-0.0243	0.0002	0.217	-0.0081^{*}
Common PCA across time	29	0.0310	-0.0006	0.000	-0.0157^{***}
B. Point estimates of SES gaps					
90–10 percentile	52	-0.0027	0.0000	0.491	-0.0027
75–25 percentile	77	0.0210	-0.0004	0.234	0.0210
70–30 percentile	89	-0.0048	0.0000	0.366	-0.0048

Notes: n = number of observations; α_1 and α_2 are the trend parameters in equation 2; the linear model sets α_2 to zero. TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment) assessments. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Complete estimates are found in table A2, available in the online appendix.

over time. Of course, data on characteristics of families other than education, books, and items in the home could change the picture. But within available measures of observed family differences, we see a consistent pattern of declining achievement gaps little affected by the specific SES construct.

Group Calculations versus Point Estimation of Gaps

In our main analysis, we estimate performance within a specific segment of the SES distribution by averaging scores across all students within the SES segment. Specifically, our preferred analysis calculates the gap between students in the top and bottom quartiles of the SES distribution as the difference between the average score of students whose families fall in the top SES quartile and the average score of students whose families fall in the bottom SES quartile. This comparison provides direct information about achievement at the extreme quartiles of the distribution but does not characterize the precise pattern of achievement within the extremes of the SES distribution.

An alternative approach—used, for example, in Reardon (2011), Chmielewski and Reardon (2016), Chmielewski (2019), and Hashim et al. (2020)—is to estimate differences in achievement for those at specific points in the SES distribution, such as the estimated difference in achievement between individuals exactly at the 75th percentile and the 25th percentile or at the 90th and 10th percentiles of the distribution. In terms of understanding the nature of achievement differences between children from “better off” versus “worse off” families, we do not think identification of differences between those at specific percentiles of the SES distribution provides as much information as our approach that aggregates achievement differences over a broader range of the distribution. Nonetheless, this approach has been followed in a range of studies, and it is worth examination to see whether this approach alters our results.[33]

---

[33] This point estimation approach has been necessitated in several of the prior analyses by the fact that direct measures of the achievement gap at chosen points of the SES distribution are not supported within their datasets and extrapolation is required to estimate achievement in the tails of the distribution (see the next

The underlying calculations involve estimating a separate linear or cubic regression function through all available SES data points observed for each test-subject-age-year. With this regression function, achievement is predicted at specific SES percentiles. Then, combining the data for the different test-subject-age-years, it is possible to estimate the trends in achievement gaps at these specific points.

We consider the same array of achievement gaps as in figure 3 except that the gaps are now calculated for the specific percentiles as opposed to the average achievement for individuals above and below the comparison levels. The qualitative patterns of the trends are the same as before, but neither the quadratic terms nor a simple linear trend are significantly different from zero (see table 6, panel B, and online appendix figure A4).

These estimations use the sample selection rules previously applied. To be included in the trend estimation, it is necessary to observe achievement for SES data points at or above the top percentile and at or below the bottom percentile. In other words, there has to be support for the SES–achievement function within the data. Thus, we use the same 77 observations for the 75–25 point estimates, the same 89 observations for the 70–30 point estimates, and the same 52 observations for the 90–10 point estimates that we did in the main calculations.[34]

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 mistaken interpretations 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).

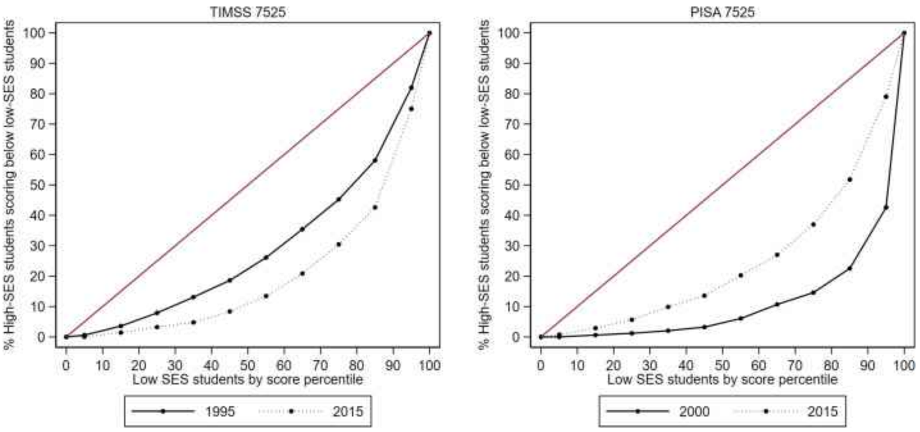
To understand better the potential impact of an ordinal approach versus the cardinal approach we previously used, we consider illustrative examples of trend analysis in SES–achievement gaps over time using only ordinal, rank-preserving assumptions of test-scale information.

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. We plot these two achievement distributions

---

footnote). Within the range of data that support observed differences in the SES–achievement relationship, questions still arise about the estimated SES–achievement relationship and the errors that are introduced into the underlying gap estimates that underlie subsequent trend analysis.

[34] The impact on trend estimation when extrapolating into the tails of the distribution is, however, much less clear. These projection problems are most severe when the SES distributions are derived from a single item where there is limited categorical data such as using parental education or categorical income (e.g., Reardon 2011; Chmielewski and Reardon 2016; Chmielewski 2019). Without information about achievement in the tails, the projection methodology can profoundly affect the estimated SES–achievement gaps. As an example, if we consider using the 2015 PISA data for parental education to derive the SES–achievement relationship, the choice of linear and cubic distributions leads to dramatically different gap estimates. The estimated 90th-percentile achievement level can differ by over 0.2 SD depending on the choice of linear or cubic projection functions. See figure A5, available in the online appendix.



Notes: Horizontal axis: students in the low-SES group (25th percentile and below ordered by their math achievement distribution. Vertical axis: share of students in the high-SES group (75th percentile and above) who score at or below the respective math achievement of the low-SES percentile. TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment.

Figure 5. Socioeconomic Status (SES) 75–25 Achievement Gaps Based on Rank Order Comparisons, TIMSS Math and PISA Math

against each other. If there was an equal achievement distribution for the top and the bottom quartile of the SES distribution, this plot would appear as a 45-degree line—just as with a Lorenz curve.[35] The greater the divergence from the 45-degree 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.

Because the analysis requires a clear identification of the two SES groups at the respective SES quartiles, we focus the analysis on the PISA and TIMSS tests again where there are relatively smooth SES distributions. We refrain from analysis of the NAEP data with this approach because of the lumpiness of the SES distributions that requires interpolation of the underlying data and thus makes precise identification of quartiles impossible.[36]

These ordinal analyses yield conclusions that follow our prior analysis. For example, using the earliest and latest installments of the TIMSS math test, the left side of figure 5 shows that the 2015 distribution is less equitable than the 2000 distribution, as the distance of the curved line from the 45-degree line for that year encloses more space than the curved line for 1995. This is just what is shown in figure 2, where the TIMSS gaps increase somewhat across birth cohorts.[37]

Using the PISA data, the right side of figure 5 shows that the 75–25 curve for 2015 has moved closer to the 45-degree line of equation. This decrease in inequality between

---

[35] 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 performance is inverted such that the share of high-SES students who score below a certain achievement threshold exceeds the equivalent share of low-SES students. These comparisons are also called probability-probability (PP) curves (Ho 2009).

[36] Ho and Reardon (2012) propose approaches for dealing with this problem of limited number of observed points in the SES distribution, although they would need to be modified for problems of interpolating among points.

[37] This pattern does, however, differ from the findings of Broer, Bai, and Fonseca (2019) for TIMSS. That study, based on a fixed and predetermined index of SES, concluded that 75–25 SES inequality declines in science but is constant in math.

2000 and 2015 is what was implied by the prior interval-based calculations shown in figure 2.

There is no easy way to combine information from multiple testing situations and regimes and to compare magnitudes, but 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.[38]

8. CONCLUSIONS

Our analysis leads to the conclusion that SES–achievement gaps have closed over the past four decades. We cannot identify the causes of this decline, but we can offer strong evidence against the current conventional wisdom that gaps have worsened (e.g., Rothstein 2004; Murray 2012; Duncan and Murnane 2014a; Putnam 2015). Our estimates indicate a slow reduction in SES–achievement gaps, but one that, if continued at the pace of the last half century, would take a very long time to erase the existing gaps. At the current closure rates, it will take almost all of the twenty-first century to cut existing SES–achievement gaps in half. Elimination of achievement gaps would not happen until past the middle of the twenty-second century.

Our analysis combines achievement data from four large-scale assessments that are designed to provide valid and reliable information about trends in student cognitive skills. Each of these assessments also includes survey information that permits construction of an SES index for each student's family. With over 1.5 million observations on test data in math, reading, and science, we can estimate trends in inequality based on family backgrounds.

We focus on trends in the 75–25 SES–achievement gap, that is, the pattern of average achievement of those in the top quartile of the SES distribution compared to achievement of those in the bottom quartile. From the 77 test-subject-age data points, we estimate quadratic trends in gaps for students with birth years from 1961–2001.

These gaps close at 0.05 SD per decade. They are not driven by the specific test regime (LTT-NAEP, Main-NAEP, TIMSS, or PISA), the subject of the test, the age of the students (age 13 or 17), the recentness of the birth cohort, or the details of the construction of the SES measure.

It is currently not possible to offer a definitive explanation of its causes. Our findings could be consistent with a variety of combinations of demographic changes and policy shifts that have occurred over the past decades. The trend could reflect changes in both the family and school inputs for students in the top and bottom parts of the SES distribution.

Research on the family has seldom moved beyond correlational analysis between specific background indicators and achievement levels, making it difficult to specify the causal structure linking specific family inputs to learning outcomes or weighting the relative importance for learning of each input. Whether or not changes in family background are widening or narrowing achievement gaps depends heavily on which

---

[38] It is conceptually possible to compare the area between the 45-degree line and the Lorenz curve for each of the 77 data points in our main analysis, but the interpretation again depends on a number of assumptions about the underlying score distributions (Ho 2009).

family input is thought to be the most critical. If income is the determinative factor, then widening gaps might be expected, given the trend toward 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. 2018). If household structure is critical, then the growing incidence of single-parent households might contribute to widening achievement gaps (Rowe 2022). But if age of the mother at the birth of the child is decisive, then disparities may be decreasing (Duncan, Kalil, and Ziol-Guest 2017). Similarly, increases in parental education and reductions in the number of children in the household—two factors consistently associated with improved student achievement—suggest narrowing gaps over time. Improvements in nutrition, health, and general economic well-being may also be disproportionately concentrated in low-SES households. The precise impact of the changes in all of these and other family factors over the past four decades remains unclear.

A similar conundrum exists on the policy side. A long list of programs has been introduced with the intention of closing SES–achievement gaps, but their success at doing so is unclear. 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 2017); significant early childhood programs such as the federal Head Start program and parallel state-funded programs (e.g., Friedman-Krauss et al. 2018); charter school formation and expansion (Shakeel and Peterson 2020); and introduction of test-based accountability focused on performances of disadvantaged students by the No Child Left Behind Act (e.g., Hanushek and Raymond 2005; Peterson 2010; Figlio and Loeb 2011). But, other changes in school inputs may have contributed to increased achievement gaps between the top and the bottom of the SES distribution. These include increased segregation of schools by SES lines, educational investments by higher-SES parents, and personnel policies that discourage the presence of high-quality teachers in schools serving low-SES students (Duncan and Murnane 2014b).

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. Cognitive skills remain critical for the income and economic well-being of U.S. citizens. The closing of achievement gaps across the SES spectrum, however modest it has been, suggests that some progress toward greater intergenerational mobility may be forthcoming. Yet, the large remaining SES–achievement gaps indicate that issues of intergenerational mobility will not soon disappear.

ACKNOWLEDGMENTS

This is a refined and refocused version of Hanushek, Peterson, Talpey, and Woessmann (2019, 2020). Helpful comments throughout the project were received from anonymous referees, Markus Broer, Matthew Chingos, Greg Duncan, Glenn Ellison, Dan Goldhaber, John Klopfer, Magne Mogstad, Richard Murnane, Randall Reback, 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 Seminar at Stanford.

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, pp. 339–358. New York: Russell Sage Foundation.

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

Autor, David H. 2014. Skills, education, and the rise of earnings inequality among the "other 99 percent." *Science* 344(843): 843–851. 10.1126/science.1251868

Bai, Yifan, Stephanie Straus, and Markus Broer. 2021. U.S. national and state trends in educational inequality due to socioeconomic status: Evidence from the 2003–17 NAEP. AIR-NAEP Working Paper. Washington, DC: AIR.

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(5): 1468–1479. 10.1162/REST_a_00370

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, pp. 255–276. New York: Russell Sage Foundation.

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

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, pp. 77–239. Cambridge, MA: MIT Press.

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(3): 486–511. 10.1080/00221546.2018.1493669

Chmielewski, Anna K. 2019. The global increase in the socioeconomic achievement gap, 1964 to 2015. *American Sociological Review* 84(3): 517–544. 10.1177/0003122419847165

Chmielewski, Anna K., and Sean F. Reardon. 2016. Patterns of cross-national variation in the association between income and academic achievement. *AERA Open* 2(3): 1–27. 10.1177/2332858416649593

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, DC: U.S. Government Printing Office.

Corak, Miles. 2013. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives* 27(3): 79–102. 10.1257/jep.27.3.79

Cross, Christopher T. 2014. *Political education: Setting the course for state and federal policy*, 2nd ed. New York: 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(5): 1927–1956. 10.1257/aer.102.5.1927

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2019. The economics of parenting. *Annual Review of Economics* 11(1): 55–84. 10.1146/annurev-economics-080218-030156

Duncan, Greg J., Ariel Kalil, and Kathleen M. Ziol-Guest. 2017. Increasing inequality in parent incomes and children's schooling. *Demography* 54:1603–1626. 10.1007/s13524-017-0600-4

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(5): 1263–1279. 10.1037/a0023875

Duncan, Greg J., and Richard J. Murnane. 2014a. Growing income inequality threatens American education. *The Phi Delta Kappan* 95(6): 8–14. 10.1177/003172171409500603

Duncan, Greg J., and Richard J. Murnane. 2014b. *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.

Egalite, Ann J. 2016. How family background influences student achievement. *Education Next* 16(2): 70–78.

Figlio, David, and Susanna Loeb. 2011. School accountability. In *Handbook of the economics of education*, vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, pp. 383–421. Amsterdam: North Holland. 10.1016/B978-0-444-53429-3.00008-9

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*. New Brunswick, NJ: Rutgers University Press.

Gould, Eric, Avi Simhon, and Bruce A. Weinberg. 2019. Does parental quality matter? Evidence on the transmission of human capital using variation in parental influence from death, divorce, and family size. NBER Working Paper No. 25495.

Goyette, Kimberly A. 2008. College for some to college for all: Social background, occupational expectations, and educational expectations over time. *Social Science Research* 37(2): 461–484. 10.1016/j.ssresearch.2008.02.002

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, pp. 181–226. Washington, DC: Brookings Institution Press.

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(3): 23–46. 10.1257/jep.22.3.23

Hanushek, Eric A. 2001. Black-white achievement differences and governmental interventions. *American Economic Review* 91(2): 24–28. 10.1257/aer.91.2.24

Hanushek, Eric A., Babs Jacobs, Guido Schwerdt, Rolf van der Velden, Stan Vermeulen, and Simon Wiederhold. 2021. The intergenerational transmission of cognitive skills: An investigation of the causal impact of families on student outcomes. NBER Working Paper No. 29450.

Hanushek, Eric A., Jacob Light, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. 2022. *Notes on trends in the U.S. income-achievement gap*. Available http://hanushek.stanford.edu/publications/notes-trends-us-income-achievement-gap.

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 No. 25648.

Hanushek, Eric A., Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. 2020. Long-run trends in the U.S. SES–achievement gap. NBER Working Paper No. 26764.

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

Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann. 2012b. Is the U.S. catching up? International and state trends in student achievement. *Education Next* 12(3): 32–41.

Hanushek, Eric A., and Margaret E. Raymond. 2005. Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management* 24(2): 297–327. 10.1002/pam.20091

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. 10.1016/j.euroecorev.2014.10.006

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2017. Coping with change: International differences in the returns to skills. *Economics Letters* 153:15–19. 10.1016/j.econlet.2017.01.007

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, pp. 89–200. Amsterdam: North Holland. 10.1016/B978-0-444-53429-3.00002-8

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) 27(1): 4–9.

Hashim, Shirin A., Thomas J. Kane, Thomas Kelley-Kemple, Mary E. Laski, and Douglas O. Staiger. 2020. Have income-based achievement gaps widened or narrowed? NBER Working Paper Series No. 27714.

Heckman, James J., and Paul A. Lafontaine. 2010. The American high school graduation rate: Trends and levels. *Review of Economics and Statistics* 92(2): 244–262. 10.1162/rest.2010.12366

Ho, Andrew D. 2009. A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics* 34(2): 201–228. 10.3102/1076998609332755

Ho, Andrew D., and Sean F. Reardon. 2012. Estimating achievement gaps from test scores reported in ordinal proficiency categories. *Journal of Educational and Behavioral Statistics* 37(4): 489–517. 10.3102/1076998611411918

Hoff, Erika. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74(5): 1368–1378. 10.1111/1467-8624.00612

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(1): 157–218. 10.1093/qje/qjv036

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, pp. 133–162. New York: Russell Sage Foundation.

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

Kao, Grace, and Marta Tienda. 1998. Educational aspirations of minority youth. *American Journal of Education* 106(3): 349–384. 10.1086/444188

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, pp. 1–75. Cambridge, MA: MIT Press.

Ladd, Helen F., ed. 1996. *Holding schools accountable: Performance-based reform in education*. Washington, DC: The Brookings Institution.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2017. School finance reform and the distribution of student achievement: Online appendix. *American Economic Journal: Applied Economics* 10(2): 1–26. 10.1257/app.20160567

Loveless, Tom. 2017. How well are American students learning? With sections on the latest international test scores, foreign exchange students, and school suspensions. *Brown Center Report on American Education* 3(6).

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.

Majoros, Erika, Monica Rosén, Stefan Johansson, and Jan-Eric Gustafsson. 2021. Measures of long-term trends in mathematics: Linking large-scale assessments over 50 years. *Educational Assessment, Evaluation and Accountability* 33(1): 71–103. 10.1007/s11092-021-09353-z

Marks, Gary N., and Michael O'Connell. 2021. Inadequacies in the SES–achievement model: Evidence from PISA and other studies. *Review of Education* 9(3): e3293. 10.1002/rev3.3293

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(6): 305–322. 10.3102/0013189X17726282

Murnane, Richard J. 2013. U.S. high school graduation rates: Patterns and explanations. *Journal of Economic Literature* 51(2): 370–422. 10.1257/jel.51.2.370

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

National Center for Education Statistics. 2012. *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. Washington, DC: National Center for Education Statistics.

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(10): 727–757. 10.1037/h0055707

Nielsen, Eric R. 2015. The income-achievement gap and adult outcome inequality. Finance and Economics Discussion Series No. 2015-041. Washington, DC: Federal Reserve Board.

Organization 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, pp. 99–157. Dordrecht: Springer Netherlands.

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, DC: Brookings.

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

Reardon, Sean F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In *Whither opportunity? Rising inequality, schools, and children's life chances*, edited by Richard J. Murnane and Greg J. Duncan, pp. 91–116. New York: Russell Sage Foundation.

Rivkin, Steve G. 2016. Desegregation since the Coleman Report: Racial composition of schools and student learning. *Education Next* 16(2): 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, pp. 1019–1049. Amsterdam: North Holland.

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

Rowe, Ian V. 2022. *Agency*. West Conshokocken, PA: Templeton Press.

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(2): 519–578. 10.1093/qje/qjw004

Shakeel, M. Danish, and Paul E. Peterson. 2020. Changes in the performance of students in charter and district sectors of U.S. education: An analysis of nationwide trends. *Journal of School Choice* 14(4): 604–632. 10.1080/15582159.2020.1811467

Shakeel, M. Danish, and Paul E. Peterson. 2022. A half century of progress in US student achievement: Agency and Flynn effects, ethnic and SES differences. *Educational Psychology Review*. 10.1007/s10648-021-09657-y

Sirin, Selcuk R. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research* 75(3): 417–453. 10.3102/00346543075003417

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, pp. 230–342. New York: Random House.

U.S. Department of Education. 2020. *Digest of education statistics 2020*. Washington, DC: National Center for Education Statistics.

von Stumm, Sophie, Nicole Cave, and Paul Wakeling. 2022. Persistent association between family socioeconomic status and primary school performance in Britain over 95 years. *NPJ Science of Learning* 7(4). 10.1038/s41539-022-00120-3

White, Karl R. 1982. The relation between socioeconomic status and academic achievement. *Psychological Bulletin* 91(3): 461–481. 10.1037/0033-2909.91.3.461

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(2): 55–69. 10.1162/DAED_a_00077

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