Program on Education Policy and Governance Working Papers Series

A Half Century of Progress in U. S. Student Achievement: Agency and Flynn Effects; Ethnic and SES Differences

M. Danish Shakeel and Paul E. Peterson [1]

PEPG 21-01

Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg/

[1] Shakeel: University of Buckingham (U.K.); Peterson: Harvard University and Hoover Institution/Stanford University.

Program on Education Policy and Governance Working Papers Series

A Half Century of Progress in U. S. Student Achievement:

Agency and Flynn Effects; Ethnic and SES Differences

M. Danish Shakeel and Paul E. Peterson

PEPG 21-01

Harvard Kennedy School
79 JFK Street, Taubman 304 Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg/

M. Danish Shakeel (danish.shakeel@buckingham.ac.uk)*

Paul E. Peterson (ppeterso@gov.harvard.edu)

December 3, 2021

*corresponding author. Prior versions of this paper were presented at the 2020 and 2021 International School Choice and Reform Conferences, the 2021 faculty seminar at Harvard Kennedy School, and the 2021 conference of the Association for Education Finance and Policy. We received valuable feedback from Patrick Wolf, Benjamin Scafidi, Robert Maranto, Jonathan Wai, Jonathan Mills, Martin West, Anna Egalite, Sean Reardon, Benjamin Arold, and Jennifer Hochschild. Antonio Wendland provided exceptional administrative support. Michael Poor provided assistance with graphs. Kathleen Wolf provided valuable editorial review. We thank the U.S. Department of Education for providing restricted-use data for the surveys analyzed in this paper (license #001006519E). We also thank the staff at the National Center for Education Statistics (NCES) for answering several questions related to these datasets. All errors are ours.

Abstract

Principals (policymakers) disagree as to whether U. S. student performance has changed over the past half century. To inform conversations, agents administered seven million psychometrically linked tests in math (m) and reading (rd) in 160 survey waves to national probability samples of cohorts born between 1954 and 2007. Estimated change in standard deviations (sd) per decade varies by agent (m: -0.10sd to 0.27sd, rd: -0.02sd to 0.12sd). Consistent with Flynn effects, median trends show larger gains in m (0.19sd) than rd (0.04sd), though rates of progress for cohorts born since 1990 have increased in rd but slowed in m. Greater progress is shown by students tested at younger ages (m: 0.31sd, rd: 0.08sd) than when tested in middle years of schooling (m: 0.17sd, rd: 0.03sd) or toward end of schooling (m: 0.06sd, rd: 0.02sd). Young white students progress more slowly (m: 0.28sd, rd: 0.09sd) than Asian (m: 46sd, rd: 0.28sd), black (m: 0.36sd, rd: 0.19sd) and Hispanic (m: 0.29sd, rd: 0.13sd) students. These ethnic differences generally attenuate as students age. Young students in the bottom quartile of the SES distribution show greater progress than those in the top quartile (difference in m: 0.08sd, in rd: 0.15sd), but the reverse is true for older students. Moderators likely include not only changes in families and schools but also improvements in nutrition, health care, and protection from contagious diseases and environmental risks. International data suggest that subject and age differentials may be due to moderators more general than just the United States.

Keywords: Agency effects, Achievement trends; Flynn Effect; LTT; NAEP; TIMSS; PIRLS; PISA; ethnicity, SES.

Principals require agents whenever they lack sufficient time or expertise to pursue their objectives directly (Bendor & Meirowitz, 2004; Epstein & O'Halloran, 1994; Gailmard, 2014; Holmstrom, 1984; Huber & Shipan, 2002; Volden, 2002). But when principals rely upon agents, they risk inefficiencies due to adverse selection and moral hazard, the shirking and diversion of resources to purposes agents prefer (Hayne & Salterio, 2014). Risks are especially large when principals do not rank objectives clearly or disagree as to their relative importance (Bernheim & Whinston, 1986; Laffont & Martimort, 2002; Laffont & Tirole, 1993; Olsen, 2014; Papadopoulos, 2014, p. 281), a not infrequent occurrence in a federal system where power is shared among principals and boundary organizations are used to limit undue political influence over scientific research. To minimize them, principals may use multiple agents (Braun & Guston, 2003).

We use the principal-agent model to interpret the efforts of policy makers to elicit information about the rate of educational progress by U. S. student cohorts. To obtain the information, principals employ agents who have the resources and skills needed to measure repeatedly the performance of representative samples of the school-age population. Since 1971, four agents have administered 160 waves of 17 psychometrically linked test surveys of student achievement in math and reading to national probability samples of U. S. student cohorts at various ages.[1] The surveys have received a good deal of public and scholarly attention, but prior research has not given systematic consideration to the range of estimates provided by agents who have varied purposes, survey designs and test content. We use individual level restricted-use student data available from the U. S. National Center for Education Statistics (NCES) to trace the progress in math and reading achievement by cohorts of U. S. students identified by four

[1] See Online Appendix A.1 for details of surveys, waves, age/grade groups.

agencies. Although agents report estimates of student progress per decade that range in math (m) from: -0.10 standard deviations (sd) to 0.27sd and in reading (rd) from -0.02sd to 0.12sd, median estimates show upward trends in both subjects regardless of age tested. However, one agent reports a downward trend in the performance of students at age 15.

Since principals do not have consistent preferences for a particular agent, we avoid giving undue emphasis to outliers by estimating true trends with median rather than mean estimates across surveys. To identify potential moderators, we compare our findings to estimates of trends in intelligence known as the "Flynn effects" literature (Hernnstein & Murray, 1994, p.307).[2] Recent meta-analyses of Flynn effects find greater growth in fluid reasoning (ability to analyze abstract relationships), than in crystallized knowledge (understanding the empirical world) (Pietschnig & Voracek, 2015; Trahan et al., 2014). We find similar differences between math and reading trends. The median upward trend by cohorts of students is steeper in m (0.19sd) than rd (0.04 sd). Cohorts of 4th grade students in other countries have also made more progress in math than in reading, suggesting common moderators are at work. Also consistent with Flynn effect estimates, we find steeper upward trends for students when tested at a younger age and a flattening in the upward math trend for the most recent period.

We observe heterogeneities. Young boys are making slightly more progress than young girls, but gender differences disappear when students are tested at an older age. Young white students progress more slowly (m: 0.28sd, rd: 0.09sd) than Asian (m: 46sd rd: 0.28sd), black (m: 0.36sd, rd: 0.19sd) and Hispanic (m: 0.29sd, rd: 0.13sd) ones.[3] Except for Asians, these

[2] Flynn effects are named after James Flynn, a New Zealand political scientist who compared intelligence in the 1930s with intelligence several decades later. Lynn (2013) shows that prior research detected rising levels of intelligence before Flynn reported his findings, and he makes a convincing case the trend be named after Runquist (1936). The conventional nomenclature remains, however.

[3] The definition of ethnicity used here depends heavily upon the definitions employed by the four survey agents, which vary slightly. See Online Appendix Table A.4 for a comparison of ethnicity definitions by agency.

differences attenuate when students are tested at an older age. The advance for those from households in the lowest quartile of the socio-economic status (SES) distribution is steeper than for those in the highest quartile among those tested at a younger age (difference in m: 0.08sd, in rd: 0.15sd). That advantage attenuates and reverses itself at older ages for white, black, and, to a lesser extent, Hispanic students, but not for Asian ones. Moderators that account for overall trends remain uncertain, but greater gains in math than in reading, especially at younger ages, could be due to improvements in neo-natal and early childhood health and well-being. Family-school interactions may account for differential trends by ethnicity and SES.

The remainder of the paper is organized in the following fashion: 1) the principal-agent problem in estimations of student achievement trends; 2) review of prior research on trends in education studies of student performance and psychometric studies of intelligence; 3) data collection and organization; 4) methodology; 5) results; and 6) discussion.

The Principal-Agent Problem in Estimation of Student Achievement Trends

Principals (policy makers) have regularly expressed concern about—and the desire to measure--rates of educational progress made by U. S. students at various ages and from differing backgrounds. The U. S. National Commission on Educational Excellence (NCEE) claims that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people (NCEE, 1983, p. 5)." According to Hirsch (1987, p. 7), the National Assessment of Educational Progress (NAEP) provides persuasive “evidence for the decline in shared knowledge.” That claim has been echoed in books with such disturbing titles as *The Literacy Hoax* (Copperman (1979), *Dumbing Down our Kids* (Sykes, 1995), *The Decline of Intelligence in America* (Itzkoff, 1994), and *The Dumbest*

Generation (Bauerlein, 2008). The other side says schools are improving, students are making progress, and that claims of deterioration are false (Berliner & Biddle, 1996; Bracey, 1992).

These writers are addressing an important question. The well-being of the next generation—and the country as a whole—depends upon continuing progress in student math and reading achievement. Higher levels of achievement have positive impacts on college attainment, future earnings, teenage pregnancy rates, physical and mental health, political participation, and many other life outcomes (Borghans et al., 2016; Chetty, Friedman, & Rockoff, 2014). Nations that show higher average levels of student achievement enjoy steeper rates of economic growth (Hanushek & Woessmann, 2008, 2012).

Given the importance of the topic, policy makers have sought information on changes in student achievement across cohorts (Jones & Olkin, 2004; Messick, Beaton, & Lord, 1983; NAEP, 2021). Policy statements, legislative enactments, and commitment of government resources to the collection and distribution of information on student cohort trends in performance reflect the strong interest of principals in gathering information on changes in student achievement. Indeed, the U. S. Commissioner of Education was directed by a statute enacted in 1867 to "report annually on the progress of students in the United States (Mullis, 2019)," though it would take a century before earnest efforts to fulfill this mandate were undertaken. Once the project began in earnest, principals confronted challenges which may be conceptualized as instances of the principal-agent problem (Bovens, Goodin, & Schillemans, 2014). Lacking the requisite technical skills, principals select one or more agents (survey organizations) with the technical capacity to gather the data (Braun & Guston, 2003). The assignment of the task to an agent creates moral hazards, as agents may shirk certain tasks or pursue goals that differ from those of principals. Also, agents may have difficulty interpreting the preferences of principals, especially when they have multiple unranked, not necessarily consistent objectives. Further,

principals, to broaden political support, may express goals and preferences in ambiguous terms (Shepsle, 1972). To reach consensus, objectives may be articulated in vague language that leaves specific goals unranked. Agents with superior access to technical information may choose to comply with ambiguously stated objectives in ways consistent with their own preferences. Multiple agents may thus take different approaches when gathering information.

Multiple principals are likely in federal systems. Shared power requires the consent and co-operation among tiers of government, such as a federal department, state agencies and local school districts. When government is tiered, actors can be both principals and agents. For example, a federal bureau (say, for example, U. S. Department of Education) may serve as the agent of policy makers (Congress and the President), but the bureau may then serve as the principal when asking lower-level agents (policy making committees, state and local governments, professional survey organizations) to act on its behalf. In scientific research, "boundary organizations," acting as both agents for policy-makers and principals for research organizations, are often established to minimize undue political influence and stabilize the relationship between the policy-making principals and scientific research (Van der Meulen, 1998). Their task is to "internalize the provisional and ambiguous character of the boundary" between policy choice and scientific execution (Guston, 2000, p. 29). But if boundary organizations are broadly representative of diverse institutions and interests, they provide an opportunity for principals to introduce new, unranked objectives (Shove, 2003). Should this occur, say Braun & Guston (2003, p. 307), "the means of principals' influence over scientists … is extremely limited."

In the case at hand, principals in the United States seek information on changes in the performance of student cohorts over time. But some of them have other objectives as well,

including comparison of performances of U. S. students with those in other countries, assessment of performance relative to expected performance on curriculum taught by schools, and information on student preparation for entering society toward the end of schooling. These objectives, though inter-related, are not identical and at times may be inconsistent. To the extent that agencies focus on alternative objectives, the data they collect may show varied estimates of performance trends.

Agents

Four agents have administered 160 waves of 17 temporally linked surveys of achievement in math and/or reading to nationally representative cohorts of U. S. students for various portions of the past half century: 1) the Long-Term Trend (LTT) version of NAEP, 2) the Main version of NAEP, 3) The International Association for the Evaluation of Educational Achievement (IEA) which administers Trends in International Math and Science Survey (TIMSS) and Progress in International Reading Literacy Study (PIRLS), and 4) the Program on Individual Student Assessment (PISA). The 17 temporally linked surveys are in math and/or reading for ages 9, 13, and 17 in LTT, grades 4, 8 and 12 in NAEP, grades 4 and 8 in TIMSS, grade 4 in PIRLS, and age 15 in PISA (For a list of surveys, waves, and age/grade groups, see Online Appendix, Table A.1). Each survey has specific purposes, sampling frames, and test content that differentiates it from the others. Several agents administer waves of tests in multiple subjects, but the largest number of tests are available for m and rd. We limit this analysis to those two subjects to obtain the best available data on change over time and consistency of results by agent. We use data on trends in achievement in other countries to explore possible moderators.

The National Center for Education Statistics (NCES) within the U. S. Department of Education, an agent acting on behalf of Congress and the President, has acted as a principal

overseeing the collection of data for all four of these agents for much of the period.[4] NCES is a boundary organization designed to insulate the collection of education information from political influences emanating from the U. S. Secretary of Education, who is otherwise responsible for education policy and programs. Unlike the Secretary, who serves at the will of the president, the NCES commissioner is appointed by the president to a fixed six-year term.[5]

Since 1988, NCES has been responsible for contracts that structure the design, administration, data analysis, and reports of all surveys by the four agents. Given the fact that all agents report to one government bureau, it might be thought all surveys would be guided by the principals ranked objectives. But, as we shall see, NCES is not the only principal. Further, its own objectives are ambiguous. On the one hand, it seeks in all four surveys to track changes in the performance of cohorts of students. But it also tries to estimate student performance against the perceived current curriculum, to compare U. S. student performance with that of students in other countries, and to measure student preparation for entry into society at the close of secondary education. Emphasis on one of these goals rather than another could generate alternative estimates of the progress of student cohorts.

LTT

The collection of information on changes in student cohort performance in the United States begins in 1967 with a compact signed by each of the fifty U. S. states creating the Education Commission of the States (ECS), which accepted responsibility for the direction of LTT in 1967, though costs were to be borne by the U.S. Office of Education. (Gilford, 2004;

[4] NCES maintains an individual-level data base for all four surveys, which is available to qualified research centers. Prior to 1988, the Education Commission of the States was responsible for LTT. See discussion below.
[5] Senate confirmation was required until 2010.

Lehmann, 2004).[6] Unable to reach a consensus on tests that obtained information about student performance in individual states or school districts, ECS agreed to report only overall results from a nationally representative sample of students and by gender, ethnicity, and size of community (Lehmann, 2004; Messick, Beaton & Lord, 1983, pp. 1-15). In 1998 Congress shifted responsibility for overseeing LTT to the National Assessment Governing Board (NAGB), a boundary organization consisting of 26 members appointed by the U. S. Secretary of Education (Bourque, 2004). By law, NAGB must include "governors, state legislators, local and state school officials, educators and researchers, business representatives, and members of the general public (NAGB, 2021)." The legislation limits the authority of the Secretary, apparently to preclude inappropriate political influence over the collection of sensitive information, saying "in the exercise of its functions, powers, and duties, the Board shall…be independent of the Secretary and the other offices and officers of the Department of Education."[7] But as a chair of NAGB observed (Olkin, 2004, p. 259):

> "There is tension, some of it built into the current structure where, for example, NAGB is supposed to set the policy but NCES . . . manages the contract. A lot of policy is contained in the contract, so that, although NAGB can set policy, NCES need not carry it out."

Acting as boundary organizations between the U. S. Department of Education and survey agents, ECS and NAGB have overseen administration of LTT to 16 waves of nationally representative samples of cohorts born as early as 1954 in reading and 1961 in math (Tables 1,

[6] Between 1968 and 1973, the U. S. Office of Education allocated $20,372,258 to cover costs associated with the administration of the LTT (Gilford, 2004, p. 178).

[7] Section 302 of Title III of Public Law 107-110 (1988) asks the National Assessment Governing Board, appointed by the U. S. Secretary of Education, to "continue to conduct the trend assessment of academic achievement at ages 9, 13, and 17 for the purpose of maintaining data on long-term trends in reading and mathematics (NAEP, 2021)."

A.2). The most recent cohort for which information is available for this analysis was born in 2002. Because the purpose of LTT is limited to tracking nationwide trends, LTT collects only a few thousand observations in each survey wave (Table 1). Samples are drawn at 9, 13, and 17 years. Though it estimates performance for a group defined by age that remains stable over time, LTT has not always been a perfectly constant measure of student performance across survey waves. As one commentator observes, "The objectives and test items became more closely aligned with school curricula; . . . its sampling, which had included young adults and out-of-school youth, was narrowed to those in school (Stedman, 2009, p. 3)." However, others fault LTT for maintaining consistent testing frameworks that "become increasingly irrelevant by failing to reflect changes in curricula and instructional practice (Pellegrino, Jones, & Mitchell, 1999, p. 78)."

NAEP

In the same 1988 legislation that Congress shifted responsibility for the administration of LTT from ECS to NAGB, it authorized NAGB to launch other tests in m and rd to representative samples of students in every state as well as in the nation as a whole (Bourque, 2004).[8] Students are assessed on material aligned to the curriculum expected to have been taught by that grade as well as to track change in performance over time. To obtain representative samples in each state, each survey wave includes over one hundred thousand observations (Table 1).[9] To meet both

[8] Section 303 of Title III of Public Law 107-110 (1988) directs NAGB to "conduct a national assessment and collect and report assessment data, including achievement data trends, in a valid and reliable manner on student academic achievement in public and private elementary schools and secondary schools at least once every 2 years, in grades 4 and 8 in reading and mathematics." The law also asks that this be done in grade 12. These legislative mandates appear to reflect competing objectives on the part of principals, as it is not clear that one can administer an assessment at a particular grade level that is valid and reliable and at the same time reliably measure achievement trends (see discussion in Bourque, 2004).

[9] According to one estimate, "the 2011 cost of the National Assessment of Educational Progress (NAEP) was $129,000,000 (Engel & Rutkowski, 2019).

objectives, the boundary organization asks agents to design additional surveys different from LTT surveys in purpose, sampling frame, and test content. Specifically, the law establishing NAEP calls for performance measurement in grades 4, 8 and 12 on tests that adjust to changes in in the perceived curriculum (Jones & Olkin, 2004). Since the age composition of a grade varies from one test administration to the next, and perceptions of the curriculum have shifted, results may not be comparable over time (Mosher, 2004; Mullis, 2004).

Still, most NAEP tests since 1990 have been psychometrically linked from one administration to the next. Before that date, agents interpreted the objectives of principals to be mainly concerned with obtaining valid information about school performance in individual states at a particular point in time. But they were then temporally linked by using "bridging questions" administered to tests in adjacent waves. We limit this analysis to cohorts tested as of this date. However, we exclude NAEP 12th grade math tests administered during the 1990s, because these were suspended by NAEP for five years until a revised version not psychometrically linked to prior ones was administered. The decisions indicate NAGB's dissatisfaction with the reliability and validity of its earlier version of the 12th grade math test. The revised version of NAEP 12th grade math tests reflects changes in high school standards and coursework, and changes in test administration practices (Grigg, Donahue, & Dion, 2007, pp. 14-15). The revisions delinked the new NAEP waves from their predecessors. For these reasons, trends in NAEP math for students tested in grade 12 are included only for students born after 1987.

TIMSS-PIRLS

The IEA also serves as a boundary organization in that it has remained a private, non-profit agency somewhat insulated from political influence. The organization, established in the 1960s, administers tests in math and science to students in grades 4 and 8 (TIMSS) and in

literacy in grade 4 (PIRLS) in numerous countries. Similar to NAEP, the original purpose of the TIMSS survey was to compare student performance in math and science across countries against a perceived international curriculum offered at these grade levels, and the first two administrations of TIMSS were not temporally linked to one another. TIMSS was redesigned after a conference attended by President George H. W. Bush and 49 of the 50 governors was held in Charlottesville, Virginia. The participants pledged that by the year 2000 "U.S. students will be first in the world in mathematics and science achievement (Klein, 2014)." To measure progress toward that goal, the third administration of TIMSS, launched in 1995, "introduced a number of innovations that enhanced its validity and reliability," including bridging questions that allowed for tracking of trends. The changes were facilitated by additional funding "by NCES and the National Science Foundation (Mullis & Martin, 2007, p. 13)." TIMSS has since been administered at four-year intervals for cohorts born between 1981 and 2005. However, it has remained faithful to its original mission by continuing to measure performance against a perceived curriculum, as is indicated by data collections organized by grade level rather than the students' age.

To measure comprehension of a curriculum, test items are drawn from frameworks "organized around two dimensions: a content dimension specifying the content to be assessed and a cognitive dimension specifying the thinking processes to be assessed." TIMSS says that "the majority of" its items "assess students' applying and reasoning skills" (IEA, 2021). Like NAEP, the questions TIMSS asks are "knowledge oriented;" its questions are "direct and abstract." It asks students, "What do you know?" (Hutchinson & Schagen, 2007, p. 54).

For years, IEA did not administer a reading or literacy test because principals did not think valid comparisons across language groups were possible. But in 2001 NCES administered

PIRLS, a literacy test, to students in grade 4 (Kastberg et al., 2013). The principals' purposes are to assess student performance against an understanding of the literacy children should be expected to achieve by grade 4, compare performances across nations, and estimate temporal change (Mullis & Martin, 2015). Results are available for cohorts born between 1991 and 2006. Information on student ethnicity is missing for 2001 and 2016, as the restricted-use data set is not available from NCES for these years. In sum, PIRLS, like TIMSS, seeks to reconcile measurement of student comprehension of the existing curriculum in many countries and to track cohort changes in performances.

PISA

PISA surveys student performance at age 15 in m and rd every three years Cohorts analyzed in this paper were born between 1985 and 2000. NCES has fully participated in and covered the cost of PISA administration and reporting within the United States.[10] Despite criticism of PISA by some educators,[11] U. S. officials have endorsed PISA findings even when test results are disappointing.[12]

PISA is an agent of the Organization for Economic Co-operation and Development (OECD), a 37-member international organization of industrialized countries that collects statistics and information on the economic and social well-being of member countries. The

[10] "The cost of implementing PISA 2012 in the US, for example, was about $6.7 million. These costs cover things like drawing a sample, recruiting schools, administering the assessment, scoring open-ended items, and. . . providing payments to students and schools (Engel & Rutkowski, 2019)." The results of the 2006 reading survey have not been released on the grounds that the administration of the test was not reliable in that year.

[11] An international group of educators wrote an open letter to PISA's director giving an array of reasons for being "frankly concerned about the negative consequences of the PISA rankings (The Guardian, 2014)."

[12] Upon the release of PISA data in 2012, U. S. Secretary of Education Arne Duncan expressed concern that "the big picture of U. S. performance . . . is a picture of educational stagnation. . . . The "educational challenge in American is not just about poor kids in poor neighborhoods. . . . It's about many kids in many neighborhoods. The [test] results underscore that educational shortcomings in the United States are not just the problems of other people's children. . . . We're running in place, as other high performing countries start to lap us (Hanushek, Peterson, & Woessmann, 2014)."

number of principals responsible for the administration is even more extensive than for the other three surveys. The Office of the OECD's Secretary-General serves as a boundary organization that insulates the collection of survey information on student achievement from the political influence by governments of member countries. When PISA results are published, its reports state "opinions . . . and arguments,", which are "published under the responsibility of the Secretary-General of the OECD," but "do not necessarily reflect the official views of OECD member countries (OECD, 2013, p. 6)." Yet the co-operation of principals in each participating country is necessary for both data collection and usually for the recovery of costs.

PISA differs from TIMMS and NAEP in that it seeks to measure "preparation for life" rather than performance against the school's curriculum.[13] For example, PISA says that the math literacy which it assesses is "an individual's capacity to identify and understand the role that mathematics plays in the world, make well-founded judgments, and use and engage with mathematics in ways that meet one's needs as a constructive, concerned, and reflective citizen (U. S. Department of Education, 2014)." Given this purpose, PISA tests are designed along lines different from those administered by NAEP and TIMSS. In a thoughtful comparison of test items across the various tests, Hutchison and Schagen (2007, p. 254) state the difference in these words:

> PISA items are aimed at life skills while TIMSS items are more knowledge oriented. Where TIMSS questions are more direct and abstract, PISA questions are more lengthy and wordy. TIMSS asks, "What do you know?" while PISA asks "What can you do?"

[13] Stedman (2009) criticizes NAEP for an excessive focus on curriculum at the expense of an objective PISA pursues: "It is not enough for students to receive high scores on a math test," he says; "in addition, we want them to be comfortable with math and readily use it in the real world (p. 31)."

Summary

Principals who have ambiguous and not always consistent objectives track change in test performance by cohorts of students by asking four agents, which have their own interests and preferences, to administer tests to representative samples of students. Although all four agents link their surveys psychometrically in ways that allow for tracking cohort trends over time, each interprets the objectives of principals in distinctive ways that reflect the agents' own interests and purposes. Originally asked by its boundary organization not to report results for states, districts or schools, LTT focuses only on national trends. However, even LTT has modified its sampling design and test content in response to perceived curriculum changes. NAEP has been asked by NAGB, its broadly representative boundary organization, to adjust its tests to changing curricular objectives and to measure student performance in every state to facilitate implementation of federal accountability goals. TIMSS-PIRLS is directed by IEA, an international non-profit organization, which acts as a boundary organization that mediates relationships between survey agents and the governments of co-operating countries. Its objectives include both tracking performance over time and estimating student performance against international expectations of what is to be learned by a certain grade. PISA, an agent of OECD, serves as a boundary organization that mediates relationships between survey agents and the principals of participating countries. It seeks to measure student readiness for participation in the social and economic world at the age of 15 when many countries end compulsory schooling. In sum, agencies report to different principals and boundary organizations who have varying objectives. Given the differences, estimates of trends in student performance are likely to vary by agent, generating noisy data that complicate the principals' task of ascertaining true trends in student performance. It remains to be seen whether clear signals can be detected amid the buzz.

Scholarly research

Research on cohort change in student achievement is bifurcated between education research and intelligence studies by psychometricians. In this selective review, we bring together key results from two traditions that have generally ignored one another (see Table A.1 for a summary of findings).

Sociology and economics of education

Numerous studies have made extensive use of surveys administered by the four agents, but the literature has yet to compare systematically their estimated rates of cohort progress. Education research provides estimates of divides by socio-economic status (SES), gender, ethnic, and school sector, and in the process cast light on progress in student performance. Some find modest gains in student achievement (Hanushek, Peterson, & Woessmann, 2012), chiefly in math (Campbell et al., 1996). Reardon, Valentino, and Shores (2012, p. 23) reports steeper upward trends in math than in reading on the LTT. Studies on ethnic achievement gaps report substantial closing of the black-white gap, but a flat trend in the most recent period (Jencks & Phillips, 1998; Magnuson & Waldfogel, 2008). Hedges and Nowell (1998) report closing of gaps between white and non-white students over a similar period. Miller (1995) finds a closing of Hispanic-white gaps and Asian-white gaps. Nowell and Hedges (1995) find more males than females among high-scoring individuals in math, but not in reading. Others have looked at influence of family structure on changes in student achievement (Grissmer et al., 1994) and the black-white gap (Phillips, Crouse, & Ralph, 1998). Elsewhere, Reardon (2011) draws upon surveys which are not psychometrically linked and finds income achievement gaps to have widened between students from households in the top and bottom 10 percent of the income distribution. Other studies find either no clear trend, or a flat trend, or declining trends in income

or SES achievement gaps in the United States (Broer, Bai, & Fonseca, 2019; Chmielewski, 2019; Hanushek et al., 2020; Hashim et al., 2020; Hedges & Nowell, 1998; OECD, 2018). Both Hanushek et al. (2020) and Hashim et al. (2020) find steeper upward trends in academic growth for those tested at a younger age. Similarly, Bloom et al. (2008) show annual gains on seven nationally normed tests are largest in elementary grades, attenuation of gains in early adolescence, and minimal gains at grade 12. Shakeel and Peterson (2020) show greater increases in math and reading achievement on NAEP at charter schools than at district schools for those tested between 2005 and 2017. Matheny et al. (2021) report growing gaps on state tests for school districts comprised of students from contrasting social and ethnic backgrounds. Yet all these studies, each valuable in its own way, leave unanswered an over-arching question: What signals emerge from the results of all psychometrically linked surveys?

Intelligence studies

Changes in intelligence by representative samples of cohorts are known as the "Flynn effect" (Herrnstein & Murray, 1994, p. 307).[14] See Table A.1. James Flynn discerned a positive trend in the intelligence quotient (IQ) of about three points per decade among those who took the Stanford-Binet intelligence tests (Flynn, 1984). Rodgers and Wänström (2007) show a Flynn effect for math (but not for reading) in the Children of the National Longitudinal Survey of Youth (NLSYC) PIAT-Math at each age between 5 and 13. Using the same data set, Ang, Rodgers and Wänström (2010) find Flynn effects for subgroups differentiated by gender, race, maternal education, income, and locale. A Flynn effect has also been identified among high ability students who took the SAT and ACT (Wai & Putallaz, 2011). Pietschnig and Voracek (2015), hereinafter PV, conduct a meta-analysis of 219 multiple administrations of the same (or

[14] See note 2 for nomenclature and early research on trends in intelligence.

similar) intelligence tests over the period 1909-2013 to cohorts of representative individuals from similar populations throughout the world. In general, they confirm Flynn's findings of average gains in IQ, gains of approximately 3 points every decade during this period. However, they find variation by region, and they find that the size of IQ gains has diminished in recent decades in industrialized countries. They also report for the most recent decades that there are larger IQ gains in Asia and Africa than in Europe and the Americas (Argentina, Brazil, Canada, Dominica, and the United States). Similarly, Meisenberg and Woodley (2013) report higher gains on the PISA and the TIMSS tests in countries with lower test scores at the beginning of the time period, from which they conclude that intelligence gains fade as a country's level of economic and social development reaches higher levels.

Like other psychometricians, PV distinguish between two subtypes of intelligence—fluid reasoning and crystallized knowledge (Pinker (2018, pp. 240-245). Fluid reasoning is the ability to perform "reasoning-based tasks that can be solved with (virtually) no prior knowledge," while crystallized knowledge "consists of knowledge-based questions that cannot be solved by reasoning" alone (Pietschnig & Voracek, 2015, p. 284). Put another way, measures of fluid reasoning estimate the ability to perform analytical tasks (such as providing the next number in an arithmetic progression), and measures of crystallized knowledge estimate the ability to answer empirical questions (such as the name of a capital or the meaning of a word). Surprisingly, PV report that it is fluid reasoning, not crystallized knowledge, that has increased at the more rapid rate in recent decades. Between 1985 and 2013, the gains in fluid reasoning have increased by approximately 2.2 IQ points per decade, down from 4.3 IQ points between 1952 and 1985 (p. 285). Still, these recent gains in fluid reasoning, if less than previous gains, are considerably larger than the gains of 0.04 IQ points per decade in crystallized knowledge between 1987 and

2011 (p. 285), down from 3 points per decade between 1962 and 1987. Others report similar differentials in the rate of progress for these two subtypes of intelligence in Anglo-American countries (Jensen, 1998, pp. 319–320; Lynn, 2009a; Nisbett et al., 2012). Finn et al. (2014) find a Pearson correlation of 0.53 between fluid reasoning and math scores, but only a 0.36 correlation with reading performance, for a selected population of Massachusetts students. Rindermann and Thompson (2013) search for Flynn effects across waves of LTT surveys. They show mean increases in math performance of 2.37 IQ points per decade but only 0.54 points in reading, a finding that accords with PV's results that identify steeper upward trends in fluid reasoning than in crystallized knowledge.

Summary

The existing literature on student progress in math and reading is bifurcated and fragmentary, although most studies suggest that progress has been made on one or more surveys at least for some students at certain ages in some subjects during some of the period. What remains to be considered is the consistency in the estimates of progress in both subjects for students at various ages and for important subgroups. Thus, we report: 1) consistency of estimates policy makers received from agents asked to survey student progress; 2) differences in the rate of progress in math and reading; 3) variation in trends by age at which a child is tested; 4) changes in the rate of progress over time, giving special attention to the most recent period for which intelligence studies find a slowdown in IQ growth in industrialized societies; 5) heterogeneities by gender, ethnicity, and SES, as well as by SES quartile for each ethnic group.

Data

To measure cohort progress, the four agents periodically administered 160 waves of tests in math and reading to national probability samples of U. S. cohorts born between 1954 and 2007

(for the dates of each test administration, see Online Appendix Table A.2). The period each survey covers and the number of observations for each set of waves are shown in Table 1. We use the individual-student data from the NCES restricted-use data set available to qualified researchers.[15] Approximately seven million student-level observations of U. S. students and approximately four and a half million observations of non-U. S. students are used in this analysis. Other features of each survey are discussed in the principal-agent section above.

Ethnicity

The definition of ethnicity used here depends heavily upon those employed by the surveys, which vary somewhat. See Online Appendix Table A.4 for a comparison of definitions and the measures taken to standardize the definition.

The ethnic composition of those tested changes substantially over the 50-year period (Table A.3). The percentage of white students taking the LTT math test declines from 81 percent in its first wave in 1978 to 55 percent for its latest wave in 2012. The percentage of black students ticks only slightly upward (from 13 percent to 14 percent) over the period, but the percentage of Hispanic students increases dramatically from 5 percent to 23 percent, and the percentage of Asian students (including those from the Pacific islands) jumps from 1 percent to 6 percent. Despite marginal agency differences in ethnicity definitions (Table A.4), similar trends are observed across all surveys.

Parental Education

Other than in the TIMSS and PIRLS 4th grade survey waves, estimates of parents' educational attainment are available. Although students are asked about their parents' education in different ways (Table A.5), all surveys allow for classification of the educational variable into

[15]However, we use publicly available datasets for PIRLS 2001, PIRLS 2016, PISA 2018 and TIMSS 2019. The latter two are used for international comparisons.

one of four categories—whether or not the parent with the higher attainment level has: 1) a four-year college degree; 2) at least some post-secondary education, 3) no more than a high school diploma; or 4) no high school diploma. Parental education indicators are available for 15 surveys.

Possession Index

None of the agencies ask children about the annual income of the household. Instead, they inquire about various items that may or may not be available in the child's home. The possession index is a simple count of the number of items reported to be in the household. This measure provides a crude proxy for household permanent income. The correlation between the number of household possessions and household annual income is 0.36 in the Education Longitudinal Study (ELS), 2002, which contains information on both variables. Although admittedly lower than desirable, it provides a rough measure of permanent income that provides an alternative to eligibility for free or reduced lunch, an annual income indicator discussed in the Appendix. Reporting errors may be particularly large when younger students are asked about items with which they are not familiar. Also, the meaning of any home possession may vary with time. For example, a computer is rare among cohorts at the beginning of this period but common for those at the end of it. TIMSS and PISA inquire about many more possessions than do LTT and NAEP (Table A.6), so it is important to estimate relationships between possessions and achievement separately for each survey. The number of possessions also varies within surveys. For the index, we count only the items which are available across all waves of each survey. Furthermore, we measure the variables consistently across all waves within a survey. The possession index is available for 17 surveys.

SES Index

The SES index, our preferred measure of SES, is estimated by extracting the first component from a principal component analysis obtained from the possession index and the original scale ranking of the parental education categories. To obtain as consistent a measure of SES as possible within each survey, we use the same measure of parental education and the same measure of household possessions across survey waves, even though on some waves additional information is available. The SES index is available for 15 surveys.

Methodology

We use the standard deviation (sd) of initial cohorts to calculate the distance (in sd) of the test-score distribution for all subsequent cohorts tested in each survey, subject, and age or grade. For each of these distributions, we estimate trends in mean performances over time by calculating the distance (in sd) of the test score distribution for each cohort's performance in each survey, subject, and age or grade level from the means of the initial cohort observations, which are set to zero. The assessments do not administer the full test to any one student. Instead, they estimate the performance of students on the test from their performance on the section administered to them. Each assessment provides various plausible values of that performance. Our empirical models use the second plausible value from each wave of the assessments. We arbitrarily use the second plausible value from each wave of the assessments. Results are robust to use of each of the first five plausible values and for the average of the five estimates. Survey weights are applied in all estimations.

Following Reardon (2011) and Hanushek et al. (2020), we extract the performance trend with a quadratic function of the birth cohort for each year of the test administered with the following equations:

$$O^{t}_{sg\gamma i} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_{sg\gamma ti} \quad \text{, where} \qquad (1)$$

O is the achievement score for student i, by subject s, assessment g, grade/age γ, and cohort t; α is the achievement trend, and ε is the error term. The results are robust to estimations from linear and cubic functions (Figures A.7-A.9).

We estimate the overall change and its associated standard error from the fitted point estimate and standard errors of the start and end points for each model (Feiveson, 1999; Gould, 1996; Oehlert, 1992; Philips & Park, 1988). We base our estimates on the delta method that calculates the variance, standard error, and Wald test statistic (z-test) from the nonlinear transformations of the estimated parameter vector from the fitted model. We repeat the above process separately for each subgroup analysis. We carry out all analyses in STATA 16.

Results

Unless otherwise indicated, all results are reported in sd per decade, which allows for direct comparisons with the Flynn effect originally estimated to be 3 IQ points or 0.21sd per decade. To avoid giving undue weight to outliers, we report median rather than the mean estimates of rates of cohort progress when summarizing results from several surveys. When summarizing results by the age a student is tested, we group surveys as follows: younger (age 9 and 4th grade); early adolescence (ages 13-15 and 8th grade); and older (age 17 and 12th grade). In estimates of differences in trends between dichotomously defined subgroups, a positive valence indicates more rapid progress by the group generally thought to be disadvantaged. A negative valence indicates the more advantaged group is making the greater progress, not that the trend is downward.

Table 1 displays the total amount of change and the rate of change in sd per decade observed by all the waves in each survey for each subject and age or grade level. The statistical

significance of each trend line is shown in columns 5 and 9. But, since the waves for each survey differ both in number and the period covered, we show in columns 6 and 10 the change per decade in sd, which allows for more direct comparisons across surveys. Inspection of these columns reveals that, except for the PISA estimates, all have a positive sign indicating at least some progress in student achievement.

[insert Table 1 here]

Agency Effects

Principals receive varying estimates of change in cohort achievement from the four agents gathering this information. The range of estimates is 0.37sd per decade in m and 0.14sd in rd (m: -0.10sd to 0.27sd, rd: -0.02sd to 0.12sd). See Table 1. NAEP and TIMSS-PIRLS estimate the steepest upward trends. NAEP estimates that the median trend for students is 0.27sd per decade in m and 0.08sd in rd; TIMSS-PIRLS estimates that the median trend is 0.25sd in m and 0.09sd in rd (Tables 2 and 3). Note that both agents adjust tests to fit the contemporary curriculum expected to be taught. The median estimate trend provided by LTT, which apparently does not adjust questions to fit the curriculum, is a more modest 0.15sd per decade in m and just 0.03sd in rd.

It is true that agents are not always testing students at the same age. But major differences between estimates are nonetheless present even when students are tested at about the same age. PISA, which estimates student preparation for participation in the post-secondary world at age 15, estimates a sizeable negative median trend of -0.10sd in m and a marginally negative trend of -0.02sd in rd. NAEP estimates that the median trend for students in 8th grade (who are only slightly younger) is 0.27sd per decade in m and 0.12sd in rd. That is a difference of 0.37sd in m and 0.14sd in rd. Given the range of estimates among agents, it is risky to rely upon surveys by

any of them to reach conclusions about progress in student achievement. We report median effects from all linked surveys.

Math-reading differentials in the United States

The median decadal trend in m comes to 0.19sd, but it is only 0.04sd in rd (Tables 2 and 3). If extrapolated to a 50-year period, the cohort gains in m would be 0.95sd, but just 0.2sd in rd. The largest difference between the two subjects is for students in 4th grade who took the NAEP tests; change per decade is 0.39sd for m but only 0.08sd for rd. PISA observes the smallest difference between the two subjects, -0.10sd in m and -0.02sd in rd.

[insert Table 2 here]

[insert Table 3 here]

Differential trends by age

The steepest upward trend is observed among students tested in m at a younger age. For those age 9 or in the 4th grade, the decadal rate of change in this subject varies between 0.23sd on the LTT to 0.39sd on the NAEP. For those tested in early adolescence, the decadal rate of change is more modest, varying between a negative trend on the PISA (-0.10sd) to positive trends on the LTT (0.15sd), the TIMSS (0.19sd) and the NAEP (0.27sd). For older students, the progress proceeds at a rate of no more than 0.05sd on the NAEP and 0.06sd on the LTT. The international comparisons reveal a similar pattern of larger rate of progress in grade 4 than grade 8 math on the TIMSS (Tables A.8-A.9).

On rd tests, it is less clear that cohort advances vary by the age at which a student is tested. In fact, there is only modest progress detected for any age group. For those at age 9 or in grade 4, the decadal rate of progress is 0.05sd on the LTT test, 0.08sd on the NAEP test and 0.09sd on the PIRLS test. For those in early adolescence, the rate of change is about the same as

for the younger students on the LTT test (0.03sd) and the NAEP test (0.12sd), but negative on the PISA test (-0.02sd). For older students, both estimates show little change by decade— 0.03sd on the NAEP test and hardly any change at all on the LTT test. In sum, progress in rd appears to have been very modest for young students as well as for those in early adolescence and minimal among those tested at an older age.

[insert Figure 1 here]

Math-reading differentials in other countries

To see whether or not subject and age differentials may be due to moderators more general than factors limited to the United States, we track trends on TIMSS-PIRLS and PISA in other countries for which this information is available. The 4th grade TIMSS-PIRLS surveys show larger cohort advances in m (0.24sd) than in rd (0.05sd). When TIMSS-PIRLS samples are restricted to industrialized nations (OECD members), larger differences between m (0.21sd) and rd (0.02sd) trends are observed. In non-OECD countries, progress in m (0.28sd) moderately outpacing gains in rd (0.19sd).

When one looks at specific countries in which 4th grade students have taken both TIMSS and PIRLS tests, progress is with one exception always greater in m than rd. In the United Kingdom, decadal change on m is 0.36sd but only 0.06sd in rd. In New Zealand, these numbers are 0.17sd, and -0.04sd, respectively. In Iran, the numbers are 0.33sd and 0.19sd, and in Hungary, they are 0.13sd and 0.04sd. In the United States, it is 0.20sd for m, and 0.08sd for rd. The one exception is Singapore, where m progress (0.26sd) trails rd (0.33sd). It is clearly possible for a country to educate its citizens in rd at a rate that outpaces the m trend, as Singapore shows, but, in general, the differential rate of progress for the two subjects observed in the United States appears in other industrialized countries as well. TIMSS also reveals a larger rate

of progress in grade 4 than grade 8 m (Tables A.8-A.9). In sum, differential trends on TIMSS-PIRLS by subject and by age tested in most industrialized countries for which data are available are consistent with PV's analyses of differential trends in fluid reasoning and crystallized knowledge intelligence.

PISA does not reveal similar differentials between m and rd performance among 15-year-old test-takers (Tables A.8-A.11). Instead, PISA finds minor difference in trends for the two subjects within the jurisdictions it surveys (-0.04sd in m; -0.01sd in rd). That could well be a function of the greater demands on crystallized knowledge required by the PISA m test, a topic we explore in the discussion section.

Progress persistence

To see whether rates of progress have persisted over time, we divide cohorts between those born before and as of 1990. As shown in Table 4, the median difference in m trends before and as of this date comes to -0.08sd per decade, an indication of slowing progress. Yet there is considerable variation across surveys. The climb in m performances on the LTT test for students at age 9 slide from 0.22sd per decade for tests administered to cohorts born before 1990 to 0.16sd per decade for those born subsequently. But at age 13, the trend picks up from 0.13sd for those born before that date to 0.21sd for more recent cohorts. At age 17, LTT m test performance declines from 0.10sd for cohorts born before 1990 to 0.02sd later on.

NAEP m trends rise less steeply as of 1990. In the 4th grade, the trend shifts upward at a substantial rate of 0.65sd per decade prior to 1990 but slows to 0.21sd in the more recent period. In the 8th grade, the progress slips from 0.44sd per decade for those born before 1990 to 0.08sd subsequently. They also perform better on the TIMSS m test if born before (0.35sd) rather than later (0.14sd). However, the steep decline on the PISA test for those born prior to 1990 (-0.36sd)

slows among those born subsequently (-0.06sd). Summing up, math performance fails to persist at the same rate among cohorts born as of 1990 in five of the seven surveys for which a comparison can be made. The PISA m test proves to be an exception to the dominant pattern, as is cohort performance at age 13 on the LTT test.

In rd, progress accelerates for cohorts born as of 1990. The median difference in trends before and as of that date is 0.08sd per decade. Once again, surveys vary. On the LTT test administered to those at age 9, progress is greater in the more recent period (0.15sd) than before 1990 (0.01sd). The same is true for the 4th grade NAEP test: before 1990 (-0.04sd); later (0.07sd). In early adolescence, the LTT test again shows greater progress for the more recent period (before 0.01sd; later 0.22sd). The NAEP test does not (before 0.22sd; later 0.10sd). However, the PISA test points in a less negative direction (before -0.31sd; later -0.04sd). Among those nearing the end of schooling, the LTT test (before, 0.00sd; later 0.09sd) shows an increase; however, the NAEP test (before, 0.05sd; later -0.07sd) shows quite the opposite. In other words, progress persistence (or at least less retrogression) in rd is observed in five of seven surveys. The NAEP tests administered to those in 8th and 12th grades are the outliers.

[insert Table 4 here]

In summary, birth cohorts born as of 1990, as compared to those born before that date, show a median trend that is -0.08sd per decade in m but 0.08sd per decade in rd. Five of seven surveys reveal a more gradual trend upward in m after 1990, and five of the seven surveys show a steeper one in rd after that date. The pattern is consistent with research on intelligence, which suggests a moderation of growth in fluid reasoning but not necessarily any change in crystallized knowledge. Still, the number of comparisons is few, and results are not uniform across surveys (Figures A.1-A.6). Moreover, in most cases we are unable to improve model fits significantly by

using quadratic or cubic models rather than a linear one that assumes constant growth over time (Figures A.7-A.9). That implies that change has occurred steadily over the past fifty years undisturbed by the many apparent perturbations that might have disrupted the learning process.

The exceptions to this generalization are the NAEP surveys in both 4th grade and 8th grade m, which show a steep upward trend for birth cohorts born in the 1970s and 1980s, with only a modest increment subsequently. This exception would be easier to interpret were it not contradicted by the steady but less dramatic increase in the LTT throughout the entire period. Altogether, signals are uncertain, though recent progress seems to have been greater in rd than m.

Heterogeneities

Gender, ethnic, and SES heterogeneities are observed. For gender, the median trend reveals a male advantage at an early age, but that fades by the end of secondary school. For ethnicity, we detect moderately slower gains for white than for Asian, African American, and Hispanic students. For SES, we observe somewhat larger gains for the lowest quartile as compared to the top quartile of the distribution when students are tested at a younger age, but that advantage attenuates for white, African American and, to a lesser extent, for Hispanic students. That pattern is not observed for Asian students, however.

Gender

There are not large differences in performance trends by gender. Across all surveys, the median differences in the achievement advantage of males trends downward in m (-0.02sd), but in rd (0.02sd) it moves in the opposite direction (Tables A.12-A.14). The male advantage is more noticeable when students are tested at a younger age (0.03sd in m, 0.04sd in rd) than in early adolescence (zero in m, 0.01sd in rd). The male achievement advantage disappears (-0.01sd in m and zero in rd) when students are tested at an older age.

Ethnicity

In nearly all surveys, trends in performances of non-white cohorts shift upward more rapidly than those for white ones (Tables 5-7). Only small differences are observed between rd and m. The median difference in Hispanic-white and black-white progress on all m surveys is 0.11sd per decade. The median Asian-white difference is 0.09sd. Among those of elementary-school age, the median difference is largest between white and Asian (0.18sd). On the four tests given in m to students in early adolescence, the median differences in the upward trend between white students and Asian, Hispanic, and black students are 0.12sd, 0.11sd and 0.09sd per decade, respectively (Tables 2 and 3). The relative position of white students usually continues to lag in high school. On the LTT m test, the decadal change for white students (0.09sd) trails that for black (0.15sd) and Hispanic (0.16sd) students, but it outpaces the decadal change registered by Asian cohorts (0.05sd). On the NAEP m test, white progress is 0.09sd per decade, less than that of both Hispanic (0.16sd) and Asian (0.21sd) students, but greater than that of African American ones (0.07sd).

Differences in trends by ethnicity are about the same for rd as for m performances. In the former subject, the largest difference is between Asians and whites, a median difference for all surveys of 0.15sd per decade. The median white-black and white-Hispanic differences in rd are 0.09sd and 0.06sd per decade.

[insert Table 5 here]

[insert Table 6 here]

[insert Table 7 here]

[insert Figure 2 here]

SES

When cohort progress is estimated by quartiles of the SES distribution, 10 of the 15 estimates show marginally larger upward trends for those in the lowest quartile than for those in the top quartile (Table 8). Four show more progress for those from the highest SES backgrounds, and one shows no difference. The trend lines vary by the point in the life cycle when the student is tested, and the differences are more favorable for disadvantaged students tested at a younger age. For those tested in m at the youngest age, the median trend is 0.08sd larger for the bottom quartile, for those tested in early adolescence, it is 0.03sd, and for the oldest cohort, the top quartile shows more progress than the lowest quartile (-0.03sd) (Table 2). The median differential in rd is large for students tested at the youngest age (0.15), lesser for those tested in early adolescence (0.11sd) and for those tested toward the end of secondary education, the trends are more favorable for the top quartile (-0.04sd) (Table 3). The PISA test, which is an outlier, shows a differential at age 15 in favor of the lowest SES quartile of 0.24sd in m and 0.29sd in rd.

[insert Table 8 here]

Consistency of SES effects across ethnic groups

NAEP has enough observations in each wave to estimate with precision achievement progress by SES quartile for each ethnic group. We report the differences between the highest and lowest SES quartiles on NAEP tests for each ethnic group in Tables 2 and 3 (see Appendix Tables A.15 – A.18 for detailed analyses). To avoid dependence on NAEP, we also report median trends from all surveys, even though the other estimations are less precise. As elsewhere, differentials are given a positive valence if the lowest SES quartile shows greater progress than the highest quartile.

Results from both the NAEP test and from the median of all surveys vary by age. When white and African American students are tested at a young age, gains that cohorts of students made in the lowest SES quartile are substantially larger than those that the cohorts made in the highest SES quartile. But the trend toward greater achievement equality attenuates by early adolescence, and it reverses itself by the time students reach the end of secondary schooling, with the top SES quartile among the older students showing a steeper upward trend than the lowest SES quartile. Attenuation is less apparent for Hispanic students and especially so for Asian students.

The downward shift in progress made by low relative to high SES students is quite pronounced for white students. On the NAEP, the lowest SES quartile of white students tested at a younger age outpaces the top quartile in m by 0.06sd per decade, but the advantage shifts to the highest quartile (-0.01sd) in early adolescence, increasing to -0.11sd by the end of secondary education. On the median for all surveys, these numbers are 0.04sd, 0.05sd, and -0.04sd, respectively. On rd tests, the relative rate of progress by SES on the NAEP declines from 0.20sd at the younger age to 0.04sd in early adolescence, then reverses to -0.14sd for older students; the median for all surveys declines from an upward trend of 0.15sd to 0.00sd, then reverses itself to -0.10sd.

For African American students, a similar pattern is apparent. The rate of cohort progress among the lowest quartile of students tested on NAEP's m exam tested at the younger age outpaces the highest quartile by 0.14sd, but that advantage fades to 0.08sd in early adolescence and then dramatically reverses itself (-0.22sd). The median trend numbers for all surveys are less dramatic but otherwise quite similar: 0.11sd, -0.03sd, and -0.09sd, respectively. On the NAEP rd test, the relative advantage of the lowest quartile in the rate of progress once again slows as

students age: 0.15sd, 0.06sd, and -0.02sd. The median for all surveys shows a similar pattern: 0.12sd, 0.06sd and 0.01sd.

Among Hispanic students, a similar pattern is apparent on the NAEP tests but less so on median results from all surveys. Those students in the lowest quartile tested on the NAEP exam in m at a younger age show gains relative to those in the highest quartile of 0.22sd. The advantage fades to 0.10sd in early adolescence and turns negative (-0.06sd) at the end of secondary education. On the rd test, the numbers are 0.28sd, 0.09sd, and 0.18sd. The median trends for all surveys do not vary as much in either m (0.10sd, 0.03sd, and 0.01sd) or in rd (0.17sd, 0.17sd, and 0.11sd).

The progress shown by the lowest SES quartile among Asian students shows a somewhat different pattern than for white and African American students. On the NAEP test, students at a younger age show gains in m registered by the lowest quartile that exceed the top quartile by 0.08sd and in rd 0.24sd per decade. In early adolescence, the lowest quartile trails the top quartile in m (-0.03sd), but exceeds it in rd (0.12sd). Among older students, the lowest SES quartile continues to outpace the highest quartile on both m (0.02sd) and rd (0.12sd) tests. Median trends from all surveys in both subjects show greater growth for the lowest quartile across all three age groups—in m 0.19sd, 0.08sd, and 0.09sd, in rd 0.06sd, 0.11sd, and 0.12sd, respectively. In sum, there is little evidence of a downward shift in the rate of progress made as low-SES (relative to high SES) Asian students as they age.

Consistency of effects across different SES indicators

As mentioned, the SES index is constructed from student reports of parental education and possessions in the home. To ascertain whether we obtain similar results with alternative SES indicators, we estimate trends separately by parental education and by possessions in the home.

In addition, we estimate SES trends by eligibility for free and reduced lunch, and family structure. For all of these alternative ways of measuring SES, we generally see more progress for students from less advantaged backgrounds than from those who are more advantaged (see text in Appendix, Tables A.7 and A.19 – A.27 for details).

Discussion

Our discussion proceeds as follows: 1) summary; 2) study limitations; 3) m-rd differentials; 4) agency effects; 5) age differentials; 6) ethnic differences; 7) SES differences; and 8) trend persistence.

Summary

Despite agency effects, many findings are robust to most estimations. Most important, progress in student achievement by cohorts of students in the United States is observed in 15 of 17 surveys. Only the two PISA surveys suggest retrogression. Much greater progress is observed in m (median estimate: 0.19sd) than in rd (0.04sd). A similar differential is observed in other countries for 4th grade performance on TIMSS m and PIRLS rd tests.

Upward trends in m are steeper for students tested at a younger age (0.31sd) than for those tested in the middle years (0.17sd) or toward the end of high school (0.06sd), but rd gains, always less than m gains, do not vary as much by age tested (0.08sd, 0.03sd, 0.02sd for the three categories, respectively). Heterogeneous effects by gender are small. By comparison, a number of ethnic heterogeneities are sizeable. The trend in both rd and m is more steeply upward for cohorts of nonwhite students than for white ones. The largest group difference is the median trend disparity between Asian and white cohorts (0.09sd in m and 0.15sd in rd). The median differentials between Hispanic and white cohorts are 0.11sd in m and 0.06sd in rd. The black-white differentials are 0.11sd in m and 0.09sd in rd. When tested at an earlier age, the median

upward trend is greater for Asian, white, black, and Hispanic students in the lowest SES categories than for those in the highest ones, but for all but Asians that difference fades as students age. For the other three ethnic groups, the highest SES quartile shows a greater rate of progress when students are tested as they are about to leave secondary schooling. Among Asian students, the lowest quartile shows greater progress (except for 8th grade LTT rd and NAEP m) no matter at what age the student is tested. The rate of change for cohorts born as of 1990 differs from those born earlier, but the direction of the shift depends on the subject. In m, the median estimate indicates a recent flattening of the upward trend (-0.08sd), but in rd a steeper upward trend (0.08) is registered.

Limitations

This descriptive study is unable to estimate causal effects. However, it does address an important question—how much progress in math and reading have cohorts of U. S. students made over the past half century? The question is of interest to principals for the very good reason that the rate of student progress is critical both for future generations and for the nation's well-being. Results should be interpreted cautiously because large agent effects are observed. To minimize dependence on any specific data collection strategy, we focus on consistency of results across agents and, to minimize influence of outliers on the estimates, we report median—not mean—results when summarizing information across surveys.

Estimates of heterogeneities are subject to some classification measurement error. Student background characteristics are not measured in identical ways by all agents (see Tables A.4 – A.7). For survey waves conducted by each agent, we standardize education and possession indicators used in the SES index, but the meaning of categories may change with the passage of

time. For example, a high school diploma may have a different meaning at the end of the period than it does at its beginning.

Math-reading differentials

The decadal rates of change in m (median estimate: 0.19sd) than in rd (0.04sd) parallel closely those for the two major types of intelligence. PV's meta-analysis shows that in recent decades fluid reasoning intelligence has trended upward in industrialized societies at a rate of 0.15sd per decade, a much steeper rate than the 0.03sd upward trend for crystallized knowledge. Similar m-rd differentials appear in nearly all industrialized countries who participate in both the 4th grade TIMSS and PIRLS surveys, suggesting that moderators are more general than characteristics of family practices or educational systems specific to any one country.

The parallel trends observed here and by PV's meta-analysis of the two major types of intelligence is probably more than coincidental. The typical m test places greater demands on fluid reasoning intelligence than does the typical rd exam. Mathematics has two domains—*language* and *content* (Hole, Grønmo, & Onstand, 2018; Schoenfield, 1967). M language consists of definitions such as the symbols in the following equation: X =2 +2. X refers to an unknown number, = to "equal" and + to "plus." M content consists of theorems about the logical relationships of symbols to one another. One acquires crystallized knowledge as one learns the language of mathematics. But once m language is learned, fluid reasoning is used to manipulate symbols to deduce theorems from defined axioms and solve formulas on tests aligned to school curricula. However, the application of formulas and theorems to real-life problems demands crystallized knowledge of the social and physical environment that goes beyond the calculation of abstract relationships. By contrast, rd performance depends heavily on crystallized knowledge.

Research shows fluid intelligence to be more strongly associated with mathematics than reading (Peng et al., 2019; Spinath, Freudenthaler, & Neubauer, 2010; Sternberg, Kaufman, & Grigorenko, 2008), probably because items that test content require reasoning ability in that they require the manipulation of abstract rules and principles (Ackerman & Lohman, 2003; Blair et al., 2005; Geary, 2011). Items on rd tests demand comprehension of elements of the empirical world, including the meaning of words, sentences, paragraphs, and features of the external environment (Barbarin et al., 2008). A longitudinal study of pre-school children finds that emergent school vocabulary is associated with gains in verbal intelligence [crystallized knowledge], but not with gains in fluid intelligence (van Tuijl & Leseman, 2007). The study also shows emergent m skills to be associated with gains in fluid reasoning. Studies in neurobiology and brain imaging have identified a distinction between fluid and crystallized intelligence (Blair, 2006; Horn & McArdle, 2007), which show the former to be associated with the brain's prefrontal cortex. Waltz et al. (1999) finds that damage to the prefrontal cortex appears to have little effect on crystallized intelligence.

PV say the differential rate of change for the two types of intelligence could be due to improved nutrition and health care (Lynn, 2009b), disease containment (Eppig, Fincher, & Thornhill, 2010; Van Panhuis et al., 2013), and reductions in environmental risks (for example, lead poisoning (Kaufman et al., 2014) and air pollution (Chay & Greenstone, 2003)). All these improvements appear to enhance the brain's analytical capacities during prenatal and infant stages of the life cycle. For example, "fending off aversive pathogens necessitates considerable amounts of energy, thereby removing important resources from brain development in early childhood (Pietschnig & Voracek, 2015, p. 293)." If student performances on tests of m content depend more on fluid reasoning than crystallized knowledge, but tests of rd content are largely

dependent upon the acquisition of crystalized knowledge, then the greater cohort progress in m than rd may be due to factors operating in early childhood and even prior to birth when brain capacity is most malleable.

Agency effects

The four survey agents asked to track changes in student achievement are to be applauded for their ambition, objectivity, resourcefulness, and endurance. But the answer to the basic question—how much progress? —remains elusive. Agency effects are substantial in part because temporal linkage via bridging questions—items repeated across waves—is not an exact science. Even more important, agents are responding to direction from principals with multiple, unranked, not always consistent, objectives. Although principals agree that tracking trends is one of their objectives, at least some of them also want the agents to measure student performance against the perceived contemporary curriculum, to compare performances across countries, and to estimate student preparation for life beyond schooling. As a result, trend estimates are likely to vary by an amount that well exceeds that which can be attributed to random sampling variation. At early adolescence, the median rd estimate varies from -0.02sd per decade on the PISA test to 0.12sd per decade on the NAEP test, a range of 0.14sd per decade. In m, the median estimate ranges wildly from -0.10sd per decade retrogression on the PISA m test to 0.27sd on the NAEP m test, a range of 0.37sd per decade. Depending on which agency's estimate is taken as authoritative, the public commentator can either lament the downward dive in American education or celebrate extraordinary advances forward.

NAEP and TIMSS-PIRLS outliers

Two surveys—the NAEP and the TIMSS-PIRLS show the most cohort progress. Both are designed to test knowledge and skills against the contemporary curriculum. Both agencies

could be over-estimating the rate of progress if the current curriculum sets lower expectations than those set previously. In other words, both NAEP and TIMSS-PIRLS could be adjusting the difficulty of their tests downward in the process of fixing questions to a contemporary curriculum.

PISA math outlier

The negative trend in PISA m (-0.10sd per decade) stands out as an exception to the positive m trends estimated by the other agents. The result is not limited to the United States. Across the industrialized world, PISA finds hardly any gains in either m or rd even while TIMSS shows large gains in m at 4th grade. However, PIRLS finds only small gains in 4th grade rd. That the TIMMS-PIRLS findings show sharp differences between subjects in many countries, while PISA finds hardly any, suggests that PISA places relatively equal emphasis on crystallized knowledge in both its m and rd tests while TIMSS m performance does not.

Since PISA testing only began with the cohort born in 1985, it might be thought that PISA m has identified a recent downturn in performance at age 15. But other surveys, though at times showing diminished m progress for cohorts born as of 1990, do not show the same retrogression (Table 4). Another possible explanation could be the reliability of tests administered at age 15. Some have argued that contemporary high school students, even those as young as age 15, are taking low-stakes tests less seriously than cohorts did in the past (Rindermann & Thompson, 2013). As testing has become increasingly pervasive and controversial in popular discourse, older students may increasingly see PISA math as a test that can be treated less seriously than other high-stakes tests (SAT, ACT, Advanced Placement and the like). This phenomenon could generate the appearance of deterioration in student achievement. But if the cause is increasing test skepticism, it should also appear on LTT and

NAEP tests administered to still older students. It is true that cohorts of these older students do not display the same growth rate as cohorts tested at a younger age, but neither do they show the pronounced negative trend that PISA m reports.

It is more likely the PISA m exception is due to the greater demands PISA places on crystallized knowledge relative to fluid reasoning. Given PISA's focus on "preparation for life,' its tests are less focused than TIMSS' on reasoning, solving routine problems and the application of concepts to numerical problems (Dossey, McCrone, & O'Sullivan, 2006). According to Hole, Grønmo, and Onstad (2018, p. 15), "more than two-thirds of the PISA mathematics items are independent of both mathematical results (theorems) and formulas." Both they and Wu (2009a) conclude that m theory is tested more frequently by TIMSS than PISA. Wu (2009a, p. 7) says "around half of the TIMSS items are not likely to appear in the PISA test. This proportion is surprisingly high. It could mean that a large part of mathematics taught in schools is not included in the PISA test." Nohara and Goldstein (2001) finds that 97% of PISA items dealt with real life situations (items requiring knowledge of the world), whereas only 48% items in NAEP and 44% items in TIMSS are so classified. Others also report that PISA has a disproportionately higher representation of items classified as data items than do either NAEP or TIMSS (Grønmo & Olsen, 2008; Neidorf et al., 2006; Nohara & Goldstein, 2001; Wu, 2009b). Wu (2009b) says the difference in the content of PISA and TIMSS tests to be the most likely explanatory factor for differential performances on the two tests in the countries she observed. Wu (2009b, pp. 44-45) also reports "there is a very high correlation between PISA mathematics and PISA reading scores…. The overlap between document reading (e.g., graphs, charts and tables) and data interpretation in mathematics becomes blurred."

PISA trends differ from those reported by other surveys in still another important respect: It shows much less retrogression for students in the bottom than in the top quarter of the SES distribution. PISA test could be the one and only honest broker, telling the truth about the "best and the brightest" in the United States. Or it could be that the bar set for those for whom it has the highest expectations has been drifting upward. PISA could be inadvertently setting steadily higher standards for what it takes to be prepared to assume high status positions in the larger society, as demands for high performance appear to shift upward.

At the same time, PISA's desire to expand its reach across the globe could inadvertently encourage the selection of less demanding questions for those preparing for low status occupations. There is no reason to think that PISA is deliberately adjusting its tests in this manner. Still, PISA has been able, with increasing success, to enlist the participation of non-OECD countries with a high density of low SES students (NCES, 2021b). Further, PISA, more than the other agents, has cut the link between curriculum and testing, leaving the agency to imagine for itself what is needed to be ready for life at age 15. There is no direct evidence that the difficulty of PISA test questions has been drifting upward for high SES students, downward for low SES ones, but trends by SES reported by PISA are strikingly different from those observed by the other agents. Whatever the cause of its exceptional findings, the range of results across agents emphasizes the danger of ignoring agency effects when estimating trends.

Age effects

We generally observe steeper upward trends in m for students tested at a younger age than for those tested at an older one. The age differential is more pronounced in m than rd. TIMSS also shows much steeper gains in other industrialized countries for students tested in m at a younger age rather than in early adolescence. All these findings are consistent with intelligence

research that suggests gains in fluid reasoning are realized in early childhood, perhaps because of improvements in nutrition and reductions in disease and environmental hazards (Lynn, 2013; Pietschnig & Voracek, 2015). Others have found the association between fluid intelligence and academic performance decreases as the child grows older (Ackerman & Lohman, 2003; Willingham, 1974). The early stages of developing foundational skills in numerical knowledge and calculation involve fluid intelligence (Fuchs et al., 2006; Östergren & Träff, 2013). In later stages children can retrieve mathematical facts for the foundational skills from memory, thereby reducing the role of fluid intelligence (Locuniak & Jordan, 2008), though other complex mathematical tasks such as fraction and algebra which are built on foundational skills, continue to place demands upon fluid reasoning (Fuchs et al., 2012; Jordan et al., 2013).

Age differentials might be driven by improved test-taking instruction combined with better co-ordination between instruction and test items (teaching-to-the-test) among younger students. But these tests are low-stakes tests not used for any accountability purposes at the student, teacher, or school level. Nor can this explanation account for the larger gains in m relative to rd in the many industrialized countries that do not have school accountability testing. O'Keefe and Rodgers (2020) show that Flynn effects could be over-estimated in studies that compare students of different ages on age-normed tests. That can have little effect on the results reported here, since changes over time are estimated by tracking performances of cohorts born at different times at a time when they have reached reach the same age or grade level.

Rindermann and Thompson (2013) offer a variety of explanations for the fadeout of gains realized among cohorts tested at a young age by the time a cohort reaches adolescence. These potential explanations include improvements in nutrition merely speeding up child development (for example, early puberty), but not affecting final maturational achievement. However, this

explanation is inconsistent with studies of Flynn effects in adults, which have found striking increases in fluid reasoning during the period. Rindermann and Thompson also offer two school-related explanations—a teenage culture that induces a negative peer pressure in high school and lower teacher quality for students at an older age. Neither would seem to account for differential fade-out rates for m and rd, unless m teacher shortages relative to others have become increasingly severe (Dee & Goldhaber, 2017; Rumberger, 1985).

Blagg and Chingos (2016) say it is not likely the attenuation of student performance by age is due to selection effects associated with rising high school graduation rates, as those rates do not increase over most of the period. Nor are results related to changes in student demographics, as trends are similar across ethnic and SES categories. Changes in the curriculum are also unlikely, as the fadeout is detected in the LTT, a test less connected to the curriculum than either NAEP or TIMSS. Finally, they point out that there has been no change in the number of unanswered questions and other indicators of disengagement that might imply recent high school cohorts are taking the tests less seriously than their predecessors.

In sum, a variety of potential moderators may explain the fade-out of gains in cohort performance as students are tested at older ages, but the matter remains ripe for further research. However, the differential fade-out between m and rd suggests that crystallized knowledge may be increasingly significant for m performance as students proceed through school.

Ethnic differences

Differential changes within families and schools may be narrowing ethnic group differences. Disproportionate (relative to white) Asian (0.09sd, 0.15sd) Hispanic (0.11sd, 0.06sd) and black (0.11sd, 0.09sd) student rates of progress in m and rd, respectively, could be a function of differential rates of change in family backgrounds. White-minority gaps in correlates of

student achievement such as parental education, household income, and number of children, have been gradually diminishing throughout the past fifty years. (Hedges & Nowell, 1998, pp. 161-67; Grissmer, Flanagan& Williamson, 1998; Duncan, Kalil, & Ziol-Guest, 2017). School interventions have also strived to redress differentials between whites and non-whites. School desegregation (Rivkin, 2016; Rivkin & Welch, 2006; Welch & Light, 1987), class size reduction (Krueger, 2003), Head Start (Morris et al. (2018)), equity law suits (Jackson, Johnson, & Persico, 2016; Lafortune, Rothstein, & Schanzenbach, 2018), English Language Learner programming (Shin, 2018), school accountability requirements (Dee & Jacobs, 2011), and greater school choice opportunities (Cheng & Peterson, 2020) have all been interpreted as having disproportionately positive impacts on disadvantaged minorities.

SES differences

Many equal opportunity programs have focused on preschool and elementary education. It may not surprise us, then, to see greater progress on rd (0.15sd) and m (0.08sd) tests for children in the lowest as compared to the highest SES quartile when tested in elementary school. Unfortunately, the differential gains apparent in elementary school in m (0.08sd) and rd (0.15sd) fade when children enter early adolescence (0.03sd, 0.11sd) and turn negative (-0.03sd, -0.04sd) by the time students near the end of their secondary education. This pattern appears for low SES students, regardless of whether their ethnic background is white, black, or Hispanic. It is not clear whether attenuation and reversal of positive trends can be attributed to a changing peer group culture or to inadequacies within the school system, or a combination thereof. However, it is worth noting that there is no attenuation of gains by low SES Asian students relative to high SES ones as students near the end of schooling. The increasingly large expectations that Asian American families have for their children's education (Liu & Xie, 2016) and the emphasis on

selection of highly qualified Asians in recent U. S. immigration policy (Lee & Zhou, 2015) might each contribute toward more rapid rates of growth among Asians from seemingly less advantaged backgrounds.

Trend persistence

The median growth trend in m attenuates for cohorts born as of 1990 (-0.08sd) but in rd it accelerates (0.08sd). The pattern is consistent with intelligence research that finds a recent flattening of the upward trend in fluid reasoning in industrialized societies but little change in crystallized knowledge. Moderators for the slowdown in math growth may be attenuating beneficial trends in nutrition, contagious diseases, and environmental risks such as pollution and lead poisoning. Family and school factors may be serving as moderators for recent gains in reading, a subject less influenced by fluid reasoning. Parental education attainment and family incomes, strong correlates of student achievement, have risen in the more recent period. School reforms—school desegregation, school accountability, more equitable financing, English Language Learner policies, and school choice—have had their greatest impact on cohorts born as of 1990.

Final observations

Families and schools appear to have played a key role in the moderation of heterogeneities across ethnic and SES groups. The two institutions may also have facilitated more rapid gains in rd among those born as of 1990.[16] Still, the research focus on families and schools in education research may distract attention away from broader social forces that could be at least as important. Recently, school closings in response to the Covid pandemic seem to

[16] These comments are based on trend data through for cohorts tested as of 2017.

have had a negative impact on learning for an entire generation of students (Kuhfeld et al., 2020) much as children suffered educational setbacks from school closures during wars (Ichino & Winter-Ebmer, 2004), strikes (Jaume & Willén, 2019; Belot & Webbink, 2010), and weather events (Goodman, 2015; Sacerdote, 2012). Indeed, PV (p. 285) detect a slowdown in intellectual growth during World War II, a likely byproduct of both school closures and worldwide disruptions of economic and social progress. It may be that drivers of change in student achievement over the past fifty years also stem in part from broad changes in social and economic well-being. Diminished progress in m for those born later could be due to a decline in returns from improved health and nutrition in advanced industrialized societies. This does not imply that families and schools are not moderators as well. There is much to be said for improving their effectiveness, particularly for students during their adolescent years. But the sharp difference between m and rd trends, the greater gains of students at an early age, and the recent flattening of growth in m performance all suggest that broader social, economic, and physical environments are no less important. It is reasonable to infer from our research that beneficial policies for every student from the very beginning of life could have as much impact on student achievement, especially in m, as focused interventions attempted later on.

References

Ackerman, P. L., & Lohman, D. F. (2003). Education and g. In H. Nyborg (Ed.), *The scientific study of general intelligence* (pp. 275–292). Amsterdam, the Netherlands: Elsevier.

Ang, S., Rodgers, J. L., & Wänström, L. (2010). The Flynn Effect within subgroups in the US: Gender, race, income, education, and urbanization differences in the NLSY-Children data. *Intelligence*, *38*(4), 367-384.

Barbarin, O. A., Early, D., Clifford, R., Bryant, D., Frome, P., Burchinal, M., . . . Pianta, R. (2008). Parental conceptions of school readiness: Relation to ethnicity, socioeconomic, status, and children's skills. *Early Education and Development, 19,* 671–701.

Bauerlein, M. (2008). *The dumbest generation: How the digital age stupefies young Americans and jeopardizes our future.* New York, NY: Jeremy P. Tarcher/Penguin.

Belot, M., & Webbink, D. (2010). Do teacher strikes harm educational attainment of students?. *Labour*, *24*(4), 391-406.

Bendor, J. & Meirowitz, A. (2004). Spatial models of delegation. *American Political Science Review*, *98*(2), 293–310.

Berliner, D. and Biddle, B. (1996). Making molehills out of molehills: Reply to Lawrence Stedman's review of the manufactured crisis. *Education Policy Analysis Archives* 4(3).

Bernheim, D. & Whinston, M. (1986). Common Agency. *Econometrica*, *54*(4), 923–942.

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

https://www.urban.org/research/publication/varsity-blues-are-high-school-students-being-left-behind/view/full_report.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, *29*(2), 109.

Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the pre-frontal cortex. *Intelligence, 33,* 93–106.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289-328.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354-13359.

Bourque, M. L. (2004). A History of the National Assessment Governing Board. In L. V. Jones & I. Olkin (Eds.), *The Nation's Report Card: Evolution and perspectives* (pp. 201–231). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Bovens, M., Goodin, R. E., & Schillemans, T. (eds) (2014). *Oxford handbook of public accountability*. Oxford, UK: Oxford University Press.

Braun, D., & Guston, D. H. (2003). Principal-agent theory and research policy: an introduction. *Science and public policy*, *30*(5), 302-308.

Bracey, G. (1992). The second Bracey report on the condition of public education. *Phi Delta Kappan 74*(2): 104–117.

Braun, D., & Guston, D. H. (2003). Principal-agent theory and research policy: An introduction. *Science and public policy*, *30*(5), 302-308.

Broer, M., Bai, Y., & Fonseca, F. (2019). *Socioeconomic inequality and educational outcomes: Evidence from twenty years of TIMSS.* IEA Research for Education and Springer Open.

Campbell, J. R., Reese, C. M., O'Sullivan, C., & Dossey, J. A. (1996). *NAEP 1994 trends in academic progress*. Washington, DC: U.S. Department of Education.

Chay, K. Y., & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*, *118*(3), 1121-1167.

Cheng, A., & Peterson, P. E. (2020). Experimentally estimated impacts of school vouchers on educational attainments of moderately and severely disadvantaged students. (EdWorkingPaper: 20-221). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/622r-tk70

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American economic review*, *104*(9), 2633-79.

Chmielewski, A. K. (2019). The global increase in the socioeconomic achievement gap, 1964 to 2015. *American Sociological Review*, *84*(3), 517-544.

Copperman, P. (1978). *The literacy hoax: The decline of reading, writing, and learning in the public schools and what we can do about it*. New York, NY: William Morrow & Company.

Dee, T. S., & Goldhaber, D. (2017). Understanding and addressing teacher shortages in the United States. The Hamilton Project, Washington, DC: Brookings Institution.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and management*, *30*(3), 418-446.

Dossey, J.A., McCrone, S. A., and O'Sullivan, C. (2006). *Problem Solving in the PISA and TIMSS 2003 Assessments* (NCES 2007-049). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from: http://nces.ed.gov/pubsearch.

Duncan, G. J., Kalil, A., & Ziol-Guest, K. M. (2017). Increasing inequality in parent incomes and children's schooling. *Demography*, *54*(5), 1603-1626.

Engel, L. & Rutkowski, D. (2019, March 28). Is PISA Worth its Cost? Some Challenges Facing Cost-Benefit Analysis of ILSAs. Laboratory of International Assessment Studies blog series. Retrieved from: https://bit.ly/2Wp3c9S

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3801-3808.

Epstein, D. & O'Halloran, S. (1994). Administrative procedures, information, and agency discretion. *American Journal of Political Science, 38*(3), 697–722.

Feiveson, A. H. (1999). FAQ: What is the delta method and how is it used to estimate the standard error of a transformed parameter? http://www.stata.com/support/faqs/stat/deltam.html.

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., . . . Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science, 25*, 736–744.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological bulletin*, *95*(1), 29.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29 – 43.

Gailmard, S. (2014). Accountability and principal-agent theory. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *Oxford Handbook of Public Accountability* (pp. 90–105). Oxford, UK: Oxford University Press.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47,* 1539 –1552.

Gilford, D. M. (2004). NAEP and the U.S Office of Education, 1971 to 1974. In L. V. Jones & I. Olkin (Eds.), *The Nation's Report Card: Evolution and perspectives* (pp. 165–183). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Goodman, J. (2015). In defense of snow days: Students who stay home when school is in session are a much larger problem. *Education Next, 15*(3), 64-69.

Gould, W. W. (1996). crc43: Wald test of nonlinear hypotheses after model estimation. Stata Technical Bulletin 29: 2–4. Reprinted in Stata Technical Bulletin Reprints, vol. 5, pp. 15–18. College Station, TX: Stata Press.

Grigg, W., Donahue, P., and Dion, G. (2007). *The Nation's Report Card: 12th-Grade Reading and Mathematics 2005* (NCES 2007-468). U.S. Department of Education, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office.

Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the Black-White test score gap narrow in the 1970s and 1980s? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 182– 226). Washington, DC: Brookings.

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). Student Achievement and the changing American family, Santa Monica, CA: RAND. *Institute for Education and Training*.

Grønmo, L. S., & Olsen, R. V. (2008). *TIMSS versus PISA: The case of pure and applied mathematics*. Paper presented at the 2nd IEA International Research Conference, Washington, DC.

Guston, D. (2000). *Between politics and science: Assuring the integrity and productivity of research*. Cambridge, UK: Cambridge University Press.

Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2014). US students from educated families lag in international tests. *Education Next*, *14*(4), 8-18.

Hanushek, E. A., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2020). *Long-run trends in the US SES-Achievement gap* (No. w26764). National Bureau of Economic Research.

Hanushek, E., Peterson, P. E., & Woessman, L. (2012). Achievement growth: International and U.S. state trends in student performance. PEPG Report 12–03, Program on Education Policy and Governance, Harvard University.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of economic growth*, *17*(4), 267-321.

Hashim, S. A., Kane, T. J., Kelley-Temple, T., Laski, M. E., & Staiger, D. O. (2020). *Have income-based achievement gaps widened or narrowed?* (No. w27714). National Bureau of Economic Research.

Hayne, C. & Salterio, S. E. (2014). Accounting and auditing. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *Oxford Handbook of Public Accountability* (pp. 421–440). Oxford, UK: Oxford University Press.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*(5220), 41-45.

Hedges, L.V., Nowell, A. (1998). Black-White test score convergence since 1965. In Jencks, C., Phillips, M. (Eds.), *The Black-White test score gap* (pp. 149–181). Washington, DC: Brookings Institution Press.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.

Hirsch, E. D. (1987). *Cultural literacy.* Boston, MA: Houghton Mifflin Company.

Hole, A., Grønmo, L. S., & Onstad, T. (2018). The dependence on mathematical theory in TIMSS, PISA and TIMSS Advanced test items and its relation to student achievement. *Large-scale Assessments in Education*, *6*(1), 1-17.

Holmström, B. (1984). On the theory of delegation. In M. Boyer & R. Kihlstrom (Eds.), *Bayesian Models in Economic Theory* (pp. 115–41). New York, NY: North-Holland.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. MacCallum (Eds.), *Factor analysis at 100 years* (pp. 205–247). Mahwah, NJ: Erlbaum.

Huber, J. & Shipan, C. (2002). *Deliberate discretion? The institutional foundations of bureaucratic autonomy*. Cambridge, UK: Cambridge University Press.

Hutchison, D., & Schagen, I. (2007). Comparisons between PISA and TIMSS—Are we the man with two watches? In T. Loveless (Ed.), *Lessons Learned: What international assessments tell us about math achievement* (pp. 227–262). Washington, DC: Brookings Institute Press.

Ichino, A., & Winter-Ebmer, R. (2004). The long-run educational cost of World War II. *Journal of Labor Economics*, *22*(1), 57-87.

IEA (2021). About TIMSS 2015. Retrieved from: http://timss2015.org/timss-2015/about-timss-2015/

Itzkoff, S. W. (1994). *The decline of intelligence in America: A strategy for national renewal*. Westport, CT: Praeger.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, *131*(1), 157-218.

Jaume, D., & Willén, A. (2019). The long-run effects of teacher strikes: Evidence from Argentina. *Journal of Labor Economics*, *37*(4), 1097-1139.

Jencks, C. & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jones, L. V., & Olkin, I. (2004). *The nation's report card: Evolution and perspectives*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116,* 45–58.

Kastberg, D., Roey, S., Ferraro, D., Lemanski, N., and Erberber, E. (2013). *U.S. TIMSS and PIRLS 2011 Technical Report and User's Guide.* (NCES 2013-046). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Kaufman, A. S., Zhou, X., Reynolds, M. R., Kaufman, N. L., Green, G. P., & Weiss, L. G. (2014). The possible societal impact of the decrease in US blood lead levels on adult IQ. *Environmental research*, *132*, 413-420.

Klein, A. (2014, September 23). Historic Summit Fueled Push for K-12 Standards, *Education Week*. Retrieved from: https://www.edweek.org/teaching-learning/historic-summit-fueled-push-for-k-12-standards/2014/09

Krueger, A. (2003). Economic considerations and class size. *Economic Journal, 113*,34–63.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher*, *49*(8), 549-565.

Laffont, J. J. & Martimort, D. (2002). *The theory of incentives: The principal-agent Model*. Princeton, NJ: Princeton University Press.

Laffont, J. J. & Tirole, J. (1993). *A theory of incentives in procurement and regulation*. Cambridge, MA, USA: MIT press.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1-26.

Lee, J., & Zhou, M. (2015). *The Asian American achievement paradox*. New York, NY: Russell Sage Foundation.

Lehmann, I. (2004). The genesis of NAEP. In L. V. Jones & I. Olkin (Eds.), *The Nation's Report Card: Evolution and perspectives* (pp. 25-92). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Liu, A., & Xie, Y. (2016). Why do Asian Americans academically outperform Whites?–The cultural explanation revisited. *Social Science Research*, *58*, 210-226.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41,* 451– 459.

Lynn, R. (2009a). Fluid intelligence but not vocabulary has increased in Britain, 1979-2008. *Intelligence*, *37*, 249–255.

Lynn, R. (2009b). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence*, *37*(1), 16-24.

Lynn, R. (2013). Who discovered the Flynn effect? A review of early studies of the secular increase of intelligence. *Intelligence*, *41*(6), 765-769.

Magnuson, K., & Waldfogel, J. (2008). *Steady gains and stalled progress: Inequality and the black-white test score gap*. New York, NY: Russell Sage.

Matheny, K. T., Thompson, M. E., Flores, C. T. & Reardon, S. F. (2021). Uneven progress: Recent trends in academic performance among U.S. school districts. Retrieved from: https://edopportunity.org/papers/seda%20district%20trends%20paper.pdf

Meisenberg, G., & Woodley, M. A. (2013). Are cognitive differences between countries diminishing? Evidence from TIMSS and PISA. *Intelligence*, *41*(6), 808-816.

Miller, L. S. (1995). *An American imperative: Accelerating minority educational advancement*. Yale University Press.

Messick, S. J., Beaton, A. E., & Lord, F. M. (1983). *NAEP reconsidered: A new design for a new era* (NAEP Report No. 83-1). Princeton, NJ: Educational Testing Service.

Morris, P. A., Connors, M., Friedman-Krauss, A., McCoy, D. C., Weiland, C., Feller, A., ... & Yoshikawa, H. (2018). New findings on impact variation from the Head Start Impact Study: Informing the scale-up of early childhood programs. *AERA Open*, *4*(2), 2332858418769287.

Mosher, F. A. (2004). What NAEP really could do. In L. V. Jones and I. Olkin (Eds.) *The nation's report card: evolution and perspectives.* (pp. 329-340). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Mullis, I. V. S. (2004). Assessing writing and mathematics. In L. V. Jones & I. Olkin (Eds.), *The Nation's Report Card: Evolution and perspectives* (pp. 361–379). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Mullis, I. V. S. (2019). *White paper on 50 Years of NAEP use: Where NAEP has been and where it should go next.* NAEP Validity Studies Panel. Retrieved from: https://www.air.org/sites/default/files/2021-06/50-Years-of-NAEP-Use-June-2019.pdf

Mullis, I. V. S., & Martin, M. O. (2007). TIMSS in perspective: Lessons learned from IEA's four decades of international mathematics assessments. In T. Loveless (Ed.), *Lessons learned:*

What international assessments tell us about math achievement (pp. 9–36). Washington, DC: Brookings Institution Press.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/pirls2016/framework.html

National Assessment of Educational Progress (2021). The NAEP Law: National Assessment of Educational Progress Authorization Act. Retrieved from: https://www.nagb.gov/about-naep/the-naep-law.html

National Assessment Governing Board (2021). Board members. Retrieved from: https://www.nagb.gov/about-naep/the-naep-law.html

National Center for Education Statistics. (2021a). NCES handbook of survey methods technical report. *National Center for Education Statistics*. Retrieved from: https://nces.ed.gov/statprog/handbook/pisa_keyconcepts.asp

National Center for Education Statistics. (2021b). Participation in PISA by year. Retrieved from: https://nces.ed.gov/surveys/pisa/countries.asp

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing mathematics content in the national assessment of educational progress (NAEP); Trends in international mathematics and science study (TIMSS); And program for international student*

assessment (PISA) 2003 assessments (NCES 2006-029). US Department of Education. Washington, DC: National Center for Education Statistics.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159.

Nohara, D., & Goldstein, A. A. (2001). *A comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study-Repeat (TIMSS-R), and the Program for International Student Assessment (PISA)* (NCES 2001-07). U.S. Department of Education, National Center for Education Statistics, Washington, DC.

OECD (2013). Foreword and Acknowledgements. In *PISA 2012 Results: What Students Know and Can Do (Volume I): Student Performance in Mathematics, Reading and Science*, OECD Publishing, Paris. Retrieved from: https://www.oecd-ilibrary.org/education/pisa-2012-results-what-students-know-and-can-do-volume-i/foreword-and-acknowledgements_9789264201118-1-en

OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. Paris, France: OECD Publishing.

Oehlert, G. W. (1992). A note on the delta method. *American Statistician, 46*(1): 27–29.

O'Keefe, P., & Rodgers, J. L. (2020). The Flynn effect can become embedded in tests: How cross-sectional age norms can corrupt longitudinal research. *Intelligence*, *82*, 101481.

Olkin, I. (2004). An interview with Chester E. Finn Jr. In L. V. Jones & I. Olkin (Eds.), *The Nation's Report Card: Evolution and perspectives* (pp. 251–265). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Olsen, J. P. (2014). Accountability and ambiguity. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *Oxford Handbook of Public Accountability* (pp. 106–123). Oxford, UK: Oxford University Press.

Östergren, R., & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology, 115,* 405– 421.

Papadopoulos, Y. (2014). Accountability and multi-level governance. In M. Bovens, R.E. Goodin, & T. Schillemans (Eds.), *Oxford Handbook of Public Accountability* (pp. 273–288). Oxford, UK: Oxford University Press.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press.

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, *145*(2), 189.

Phillips, M. Crouse, J. & Ralph, J. (1998). Does the black-white test score gap widen after children enter school?. In Jencks, C., Phillips, M. (Eds.), *The Black-White test score gap* (pp. 229–272). Washington, DC: Brookings Institution Press.

Phillips, P. C. B., & Park. J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica 56*: 1065–1083.

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, *10*(3), 282-306.

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. New York, NY: Viking.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–115). New York, NY: Russell Sage.

Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012). Patterns of literacy among US students. *The Future of Children*, 17-37.

Rindermann, H., & Thompson, J. (2013). Ability rise in NAEP and narrowing ethnic gaps?. *Intelligence*, *41*(6), 821-831.

Rivkin, S., & Welch, F. (2006). Has school desegregation improved academic and economic outcomes for blacks?. *Handbook of the Economics of Education*, *2*, 1019-1049.

Rivkin, S. (2016). Desegregation since the Coleman report. *Education Next,16*(2), 28–37.

Rodgers, J. L., & Wänström, L. (2007). Identification of a Flynn Effect in the NLSY: Moving from the center to the boundaries. *Intelligence*, *35*(2), 187-196.

Rumberger, R. (1985). The shortage of mathematics and science teachers: A review of the evidence. *Educational Evaluation and Policy Analysis*, *7*(4), 355-369.

Runquist, E. A. (1936). Intelligence test scores and school marks in 1928 and 1933. *School & Society, 43*, 301–304.

Sacerdote, B. (2012). When the saints go marching out: Long-term outcomes for student evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, *4*(1), 109-35.

Shakeel, M. D., & Peterson, P. E. (2020). Changes in the performance of students in charter and district sectors of US education: An analysis of nationwide trends. *Journal of School Choice, 14*(4), 604-632.

Shin, N. (2018). The effects of the initial English language learner classification on students' later academic outcomes. *Educational Evaluation and Policy Analysis*, *40*(2), 175-195.

Shepsle, K. A. (1972). The strategy of ambiguity: Uncertainty and electoral competition. *American Political science review*, *66*(2), 555-568.

Shoenfield, J. R. (1967). *Mathematical Logic*. Urbana: Association for Symbolic Logic, University of Illinois.

Shove, E. (2003). Principals, agents and research programmes. *Science and Public Policy, 30*(5): 371-81.

Spinath, B., Freudenthaler, H. H., & Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences*, *48*(4), 481-486.

Stedman, L. (2009). *The NAEP long-term trend assessment: A review of its transformation, use, and findings*. Washington DC: National Assessment Governing Board.

Sternberg, R. J., Kaufman, J. C., & Grigorenko, E. L. (2008). *Applied Intelligence*. Cambridge, UK: Cambridge University Press.

Sykes, C. J. (1995). Dumbing down our kids: *Why American children feelgood about themselves but can't read, write, or add*. New York: St. Martin's Press Griffin.

The Guardian. (2014, May 6). OECD and Pisa tests are damaging education worldwide - academics, *The Guardian*. Retrieved from: https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: a meta-analysis. *Psychological bulletin*, *140*(5), 1332.

U. S. Department of Education. (2014). Program for International Student Assessment (PISA). Retrieved from: https://nces.ed.gov/statprog/handbook/pisa.asp

Van der Meulen, B. (1998). Science policies as principal–agent games: Institutionalization and path dependency in the relation between government and science. *Research policy*, *27*(4), 397-414.

Van Panhuis, W. G., Grefenstette, J., Jung, S. Y., Chok, N. S., Cross, A., Eng, H., ... & Burke, D. S. (2013). Contagious diseases in the United States from 1888 to the present. *The New England journal of medicine*, *369*(22), 2152.

van Tuijl, C., & Leseman, P. P. (2007). Increases in the verbal and fluid cognitive abilities of disadvantaged children attending preschool in the Netherlands. *Early Childhood Research Quarterly*, *22*(2), 188-203.

Volden, C. (2002). A formal model of the politics of delegation in a separation of powers system. *American Journal of Political Science*, *46*(1), 111–133.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological science*, *10*(2), 119-125.

Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, *39*(6), 443-455.

Welch, F., & Light, A. (1987). *New evidence on school desegregation*. Washington, DC: U.S. Commission on Civil Rights.

Willingham, W. W. (1974). Predicting Success in Graduate Education: Improved selection procedures are likely to come from better definitions of "success." *Science, 183,* 273–278.

Wu, M. (2009a). A critical comparison of the contents of the PISA and TIMSS mathematics assessments. Melbourne: Assessment Research Centre, University of Melbourne. Retrieved from: https://edsurveys.rti.org/PISA/documents/WuA_Critical_Comparison_of_the_Contents_of_PISA_and_TIMSS_psg_WU_06.1.pdf

Wu, M. (2009b). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects*, *39*(1), 33.

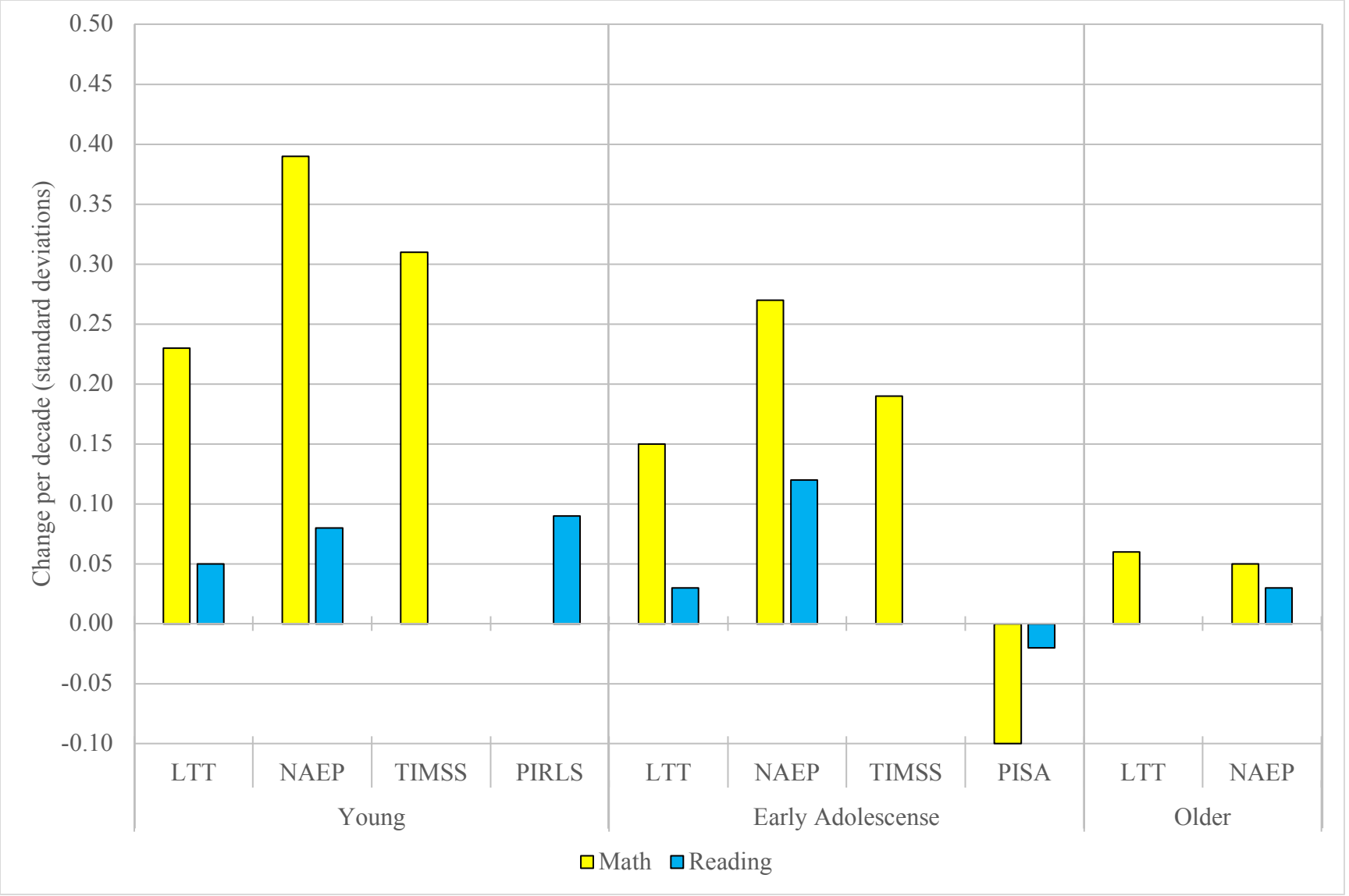

Figure 1. Change/decade in achievement levels in math and reading by age and survey: Birth Cohort 1954-2007.
Note: Figure shows change/decade in achievement levels by age and survey from the estimates in Table 1. See Tables 1, 2, and 3.
Source: See Table 1.

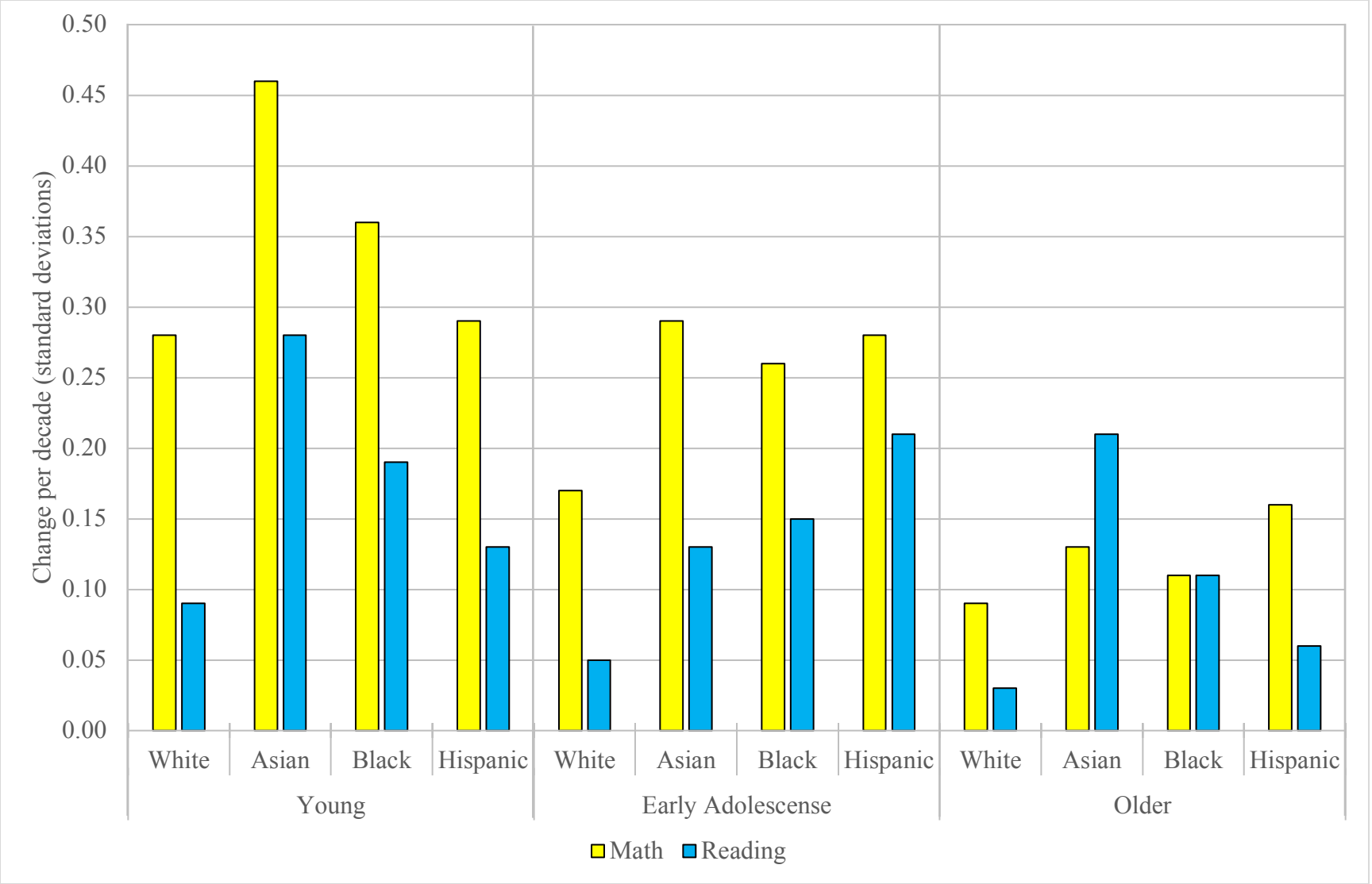

Figure 2. Median change/decade in achievement levels in math and reading by ethnicity and age: Birth Cohort 1954-2007.
Note: Figure shows median change/decade in achievement levels by ethnicity and age from the estimates in Tables 2 and 3.
See Tables 1, 2, and 3.
Source: See Tables 2 and 3.

Table 1: Change in achievement levels in math and reading by grade/age and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Survey	Grade/Age	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
LTT	Age 9	1968-2002	89,170	0.783**	0.23	1961-2002	137,750	0.222**	0.05
				(0.012)				(0.011)	
NAEP	Grade 4	1980-2007	1,455,630	1.051**	0.39	1980-2007	1,606,030	0.227**	0.08
				(0.002)				(0.002)	
TIMSS	Grade 4	1985-2005	51,440	0.610**	0.31				
				(0.011)					
PIRLS	Grade 4					1991-2006	26,100	0.129**	0.09
								(0.015)	
LTT	Age 13	1964-1998	101,200	0.493**	0.15	1957-1998	142,060	0.126**	0.03
				(0.011)				(0.012)	
NAEP	Grade 8	1976-2003	1,301,960	0.716**	0.27	1976-2003	1,423,560	0.325**	0.12
				(0.003)				(0.002)	
TIMSS	Grade 8	1981-2001	57,030	0.374**	0.19				
				(0.011)					
PISA	Age 15	1985-2000	29,130	-0.145**	-0.10	1985-2000	25,230	-0.031**	-0.02
				(0.016)				(0.016)	
LTT	Age 17	1961-1995	92,460	0.218**	0.06	1954-1995	137,350	0.013**	0.00
				(0.013)				(0.011)	
NAEP	Grade 12	1988-1998	71,500	0.048**	0.05	1973-1998	145,160	0.071**	0.03
				(0.009)				(0.008)	

Note: Table shows change in achievement levels by grade/age and survey. Normalized achievement is measured in standard deviations (s.d.). The s.d. is the difference between the year test was administered and the starting year for a specific test series. The changes have been estimated from a quadratic fit. Change per decade (dc) in achievement levels is also displayed. Standard error are in parenthesis. *p <0.05, **p<0.01.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1990-2017 Main NAEP, 1971-2012 LTT NAEP, 1995-2015 TIMSS, 2000-2015 PISA and 2001-2016 PIRLS.

Table 2: Medians of change/decade in achievement levels in math by subgroups and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Base	Subgroup	Young	Diff.	E. A.	Diff.	Older	Diff.	All age	Diff.
	All	0.31		0.17		0.06		0.19	
	LTT	0.23		0.15		0.06		0.15	
	NAEP	0.39		0.27		0.05		0.27	
	TIMSS	0.31		0.19				0.25	
	PISA			-0.10				-0.10	
Bir yr: Until 1990	Until 1990	0.44		0.24		0.10		0.22	
	Sin. 1990	0.19	-0.25	0.11	-0.13	0.02	-0.08	0.14	-0.08
Gender: Female	Female	0.29		0.17		0.06		0.20	
	Male	0.32	0.03	0.17	0.00	0.05	-0.01	0.18	-0.02
Ethnicity: White	White	0.28		0.17		0.09		0.18	
	Asian	0.46	0.18	0.29	0.12	0.13	0.04	0.27	0.09
	Black	0.36	0.08	0.26	0.09	0.11	0.02	0.29	0.11
	Hispanic	0.29	0.01	0.28	0.11	0.16	0.07	0.29	0.11
SES quart: Top	Top	0.42		0.16		0.09		0.13	
	Second	0.45	0.03	0.17	0.01	0.06	-0.03	0.17	0.04
	Third	0.44	0.02	0.19	0.03	0.03	-0.06	0.16	0.03
	Bottom	0.50	0.08	0.19	0.03	0.06	-0.03	0.16	0.03
Asian SES: Top	NAEP top	0.46		0.34		0.18		0.34	
	Bottom	0.54	0.08	0.31	-0.03	0.20	0.02	0.31	-0.03
	All top	0.24		0.21		0.09		0.14	
	Bottom	0.42	0.19	0.28	0.08	0.19	0.09	0.28	0.15
Black SES: Top	NAEP top	0.65		0.42		0.27		0.42	
	Bottom	0.79	0.14	0.50	0.08	0.05	-0.22	0.50	0.07
	All top	0.48		0.27		0.19		0.29	
	Bottom	0.59	0.11	0.24	-0.03	0.11	-0.09	0.24	-0.05
Hispan. SES: Top	NAEP top	0.64		0.37		0.22		0.37	
	Bottom	0.86	0.22	0.47	0.10	0.16	-0.06	0.47	0.10
	All top	0.42		0.26		0.18		0.21	
	Bottom	0.51	0.10	0.29	0.03	0.19	0.01	0.23	0.02
White SES: Top	NAEP top	0.64		0.40		0.17		0.40	
	Bottom	0.70	0.06	0.39	-0.01	0.06	-0.11	0.39	0.00
	All top	0.45		0.15		0.11		0.17	
	Bottom	0.50	0.04	0.20	0.05	0.07	-0.04	0.20	0.03

Notes & Source: See Table 1. Table displays medians of change/decade in standard deviations in achievement levels in math by subgroups and survey displayed in Tables 1, 4-8 and A.12 – A.18. E. A. = Early Adolescence. Bir yr = Birth year, Sin. = Since, All = LTT, NAEP, TIMSS and PISA. Birth years differ across subgroups, depending on the availability of data (see tables and appendix for details). Differences between the base category and other categories are also displayed.

Table 3: Medians of change/decade in achievement levels in reading by subgroups and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Base	Subgroup	Young	Diff.	E. A.	Diff.	Older	Diff.	All age	Diff.
	All	0.08		0.03		0.02		0.04	
	LTT	0.05		0.03		0.00		0.03	
	NAEP	0.08		0.12		0.03		0.08	
	PIRLS	0.09						0.09	
	PISA			-0.02				-0.02	
Bir yr: Until 1990	Until 1990	-0.02		0.01		0.03		0.01	
	Sin. 1990	0.11	0.13	0.10	0.09	0.01	-0.02	0.09	0.08
Gender: Female	Female	0.06		0.03		0.02		0.04	
	Male	0.10	0.04	0.04	0.01	0.02	0.00	0.06	0.02
Ethnicity: White	White	0.09		0.05		0.03		0.06	
	Asian	0.28	0.19	0.13	0.08	0.21	0.18	0.21	0.15
	Black	0.19	0.10	0.15	0.10	0.11	0.08	0.15	0.09
	Hispanic	0.13	0.04	0.21	0.16	0.06	0.03	0.12	0.06
SES quart: Top	Top	-0.06		0.04		0.06		0.00	
	Second	-0.05	0.01	0.02	-0.02	0.01	-0.05	0.00	0.00
	Third	-0.04	0.02	0.04	0.00	0.02	-0.04	0.00	0.00
	Bottom	0.09	0.15	0.15	0.11	0.02	-0.04	0.03	0.03
Asian SES: Top	NAEP top	0.02		-0.02		0.07		0.02	
	Bottom	0.26	0.24	0.10	0.12	0.19	0.12	0.16	0.14
	All top	0.03		-0.02		0.05		0.03	
	Bottom	0.09	0.06	0.10	0.11	0.17	0.12	0.16	0.13
Black SES: Top	NAEP top	0.09		0.13		0.07		0.10	
	Bottom	0.24	0.15	0.19	0.06	0.05	-0.02	0.19	0.09
	All top	0.01		0.07		0.08		0.07	
	Bottom	0.13	0.12	0.13	0.06	0.09	0.01	0.12	0.05
Hispan. SES: Top	NAEP top	-0.07		0.19		0.04		0.05	
	Bottom	0.21	0.28	0.28	0.09	0.22	0.18	0.21	0.16
	All top	0.01		0.11		0.05		0.05	
	Bottom	0.18	0.17	0.28	0.17	0.16	0.11	0.21	0.16
White SES: Top	NAEP top	-0.04		0.17		0.12		0.03	
	Bottom	0.16	0.20	0.21	0.04	-0.02	-0.14	0.16	0.13
	All top	-0.05		0.05		0.08		0.03	
	Bottom	0.10	0.15	0.05	0.00	-0.02	-0.10	0.05	0.01

Notes & Source: See Tables 1 and 2.

Table 4: Comparing the change in achievement levels in surveys until and since birth year 1990.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
		Until birthyr 1990		Since birthyr 1990		Until birthyr 1990		Since birthyr 1990	
Survey	Grade/Age	Change	Change/dc	Change	Change/dc	Change	Change/dc	Change	Change/dc
LTT	Age 9	0.466**	0.22	0.125**	0.16	0.026*	0.01	0.120**	0.15
		(0.013)		(0.015)		(0.012)		(0.014)	
NAEP	Grade 4	0.645**	0.65	0.355**	0.21	-0.037**	-0.04	0.116**	0.07
		(0.015)		(0.003)		(0.013)		(0.002)	
LTT	Age 13	0.343**	0.13	0.164**	0.21	0.043**	0.01	0.178**	0.22
		(0.013)		(0.014)		(0.013)		(0.016)	
NAEP	Grade 8	0.571**	0.44	0.100**	0.08	0.292**	0.22	0.122**	0.10
		(0.006)		(0.003)		(0.004)		(0.003)	
TIMSS	Grade 8	0.276**	0.35	0.113**	0.14				
		(0.014)		(0.013)					
PISA	Age 15	-0.108**	-0.36	-0.057**	-0.06	-0.092**	-0.31	-0.024	-0.04
		(0.023)		(0.017)		(0.021)		(0.018)	
LTT	Age 17	0.254**	0.10	0.008	0.02	0.016	0.00	0.034*	0.09
		(0.015)		(0.013)		(0.012)		(0.014)	
NAEP	Grade 12					0.073**	0.05	-0.039**	-0.07
						(0.010)		(0.008)	

Notes & Source: See Table 1. The table compares the changes in achievement levels in LTT, NAEP and TIMSS until and since birth year (birthyr) 1990 (or the closest available; see Table 1 for details).

Table 5: Change in achievement levels at younger age by ethnicity and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Ethnicity	N	Change	Change/dc	N	Change	Change/dc
LTT	Asian	2,520	0.928**	0.27	3,490	1.131**	0.31
			(0.054)			(0.044)	
	Black	14,070	0.974**	0.29	22,020	0.622**	0.15
			(0.028)			(0.025)	
	Hispanic	14,190	0.977**	0.29	14,080	0.442**	0.11
			(0.023)			(0.024)	
	White	56,420	0.853**	0.25	96,260	0.267**	0.07
			(0.015)			(0.014)	
NAEP	Asian	69,280	1.230**	0.46	76,960	0.470**	0.17
			(0.012)			(0.011)	
	Black	250,190	1.305**	0.48	278,640	0.509**	0.19
			(0.006)			(0.005)	
	Hispanic	262,040	1.286**	0.48	271,750	0.352**	0.13
			(0.006)			(0.005)	
	White	810,110	1.077**	0.40	907,840	0.255**	0.09
			(0.003)			(0.003)	
TIMSS	Asian	2,330	1.274**	0.64			
			(0.050)				
	Black	7,700	0.724**	0.36			
			(0.027)				
	Hispanic	11,530	0.589**	0.29			
			(0.023)				
	White	24,820	0.559**	0.28			
			(0.015)				
PIRLS	Asian				840	0.140*	0.28
						(0.066)	
	Black				2,410	0.233**	0.47
						(0.035)	
	Hispanic				4,500	0.160**	0.32
						(0.026)	
	White				8,980	0.168**	0.34
						(0.018)	

Notes & Source: See Table 1. See Table A.1 for details of ethnic coding.

Table 6: Change in achievement levels at early adolescence by ethnicity and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Ethnicity	N	Change	Change/dc	N	Change	Change/dc
LTT	Asian	3,150	0.806**	0.24	3,530	1.152**	0.31
			(0.046)			(0.053)	
	Black	15,880	0.739**	0.22	21,730	0.606**	0.15
			(0.026)			(0.028)	
	Hispanic	12,820	0.735**	0.22	12,300	0.147**	0.04
			(0.022)			(0.031)	
	White	68,070	0.553**	0.16	102,760	0.192**	0.05
			(0.014)			(0.014)	
NAEP	Asian	64,080	0.972**	0.36	69,690	0.349**	0.13
			(0.014)			(0.012)	
	Black	223,150	0.947**	0.35	242,490	0.411**	0.15
			(0.006)			(0.006)	
	Hispanic	213,560	0.959**	0.36	222,510	0.562**	0.21
			(0.007)			(0.006)	
	White	750,100	0.728**	0.27	833,220	0.360**	0.13
			(0.003)			(0.003)	
TIMSS	Asian	2,660	0.668**	0.33			
			(0.052)				
	Black	8,000	0.605**	0.30			
			(0.028)				
	Hispanic	11,610	0.660**	0.33			
			(0.023)				
	White	30,630	0.369**	0.18			
			(0.014)				
PISA	Asian	1,180	-0.090	-0.06	1,030	-0.084	-0.06
			(0.086)			(0.082)	
	Black	4,040	0.027**	0.02	3,580	0.057	0.04
			(0.037)			(0.038)	
	Hispanic	6,480	0.153**	0.10	5,820	0.332**	0.22
			(0.032)			(0.032)	
	White	15,340	-0.213**	-0.14	12,920	-0.101**	-0.07
			(0.021)			(0.021)	

Notes & Source: See Table 1.

Table 7: Change in achievement levels at older age by ethnicity and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Ethnicity	N	Change	Change/dc	N	Change	Change/dc
LTT	Asian	2,780	0.161**	0.05	3,410	0.840**	0.23
			(0.050)			(0.045)	
	Black	13,360	0.520**	0.15	19,200	0.689**	0.17
			(0.029)			(0.026)	
	Hispanic	9,260	0.544**	0.16	10,400	0.137**	0.03
			(0.026)			(0.029)	
	White	66,110	0.289**	0.09	103,050	0.044**	0.01
			(0.016)			(0.013)	
NAEP	Asian	3,530	0.212**	0.21	6,810	0.459**	0.18
			(0.044)			(0.039)	
	Black	10,120	0.072**	0.07	21,560	0.088**	0.04
			(0.022)			(0.020)	
	Hispanic	11,160	0.162**	0.16	21,140	0.213**	0.09
			(0.022)			(0.020)	
	White	45,470	0.089**	0.09	92,850	0.097**	0.04
			(0.011)			(0.010)	

Notes & Source: See Table 1.

Table 8: Change in achievement levels at by SES quartiles, age categories, and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Younger							
Top	LTT	46,410	0.506**	0.24	30,860	-0.131**	-0.07
			(0.030)			(0.034)	
Second			0.544**	0.26		-0.068	-0.04
			(0.030)			(0.035)	
Third			0.501**	0.24		-0.179**	-0.09
			(0.031)			(0.036)	
Bottom			0.582**	0.28		0.046	0.02
			(0.032)			(0.038)	
Top	NAEP	147,430	0.761**	0.59	161,090	-0.049**	-0.04
			(0.014)			(0.012)	
Second			0.822**	0.63		-0.060**	-0.05
			(0.014)			(0.013)	
Third			0.825**	0.63		0.027*	0.02
			(0.014)			(0.013)	
Bottom			0.930**	0.72		0.211**	0.16
			(0.014)			(0.013)	
Early Adolescence							
Top	LTT	90,010	0.454**	0.13	64,070	0.125**	0.04
			(0.022)			(0.025)	
Second			0.568**	0.17		0.078**	0.02
			(0.022)			(0.025)	
Third			0.531**	0.16		0.113**	0.04
			(0.022)			(0.025)	
Bottom			0.543**	0.16		0.098**	0.03
			(0.023)			(0.026)	
Top	NAEP	587,020	0.709**	0.37	703,770	0.283**	0.15
			(0.007)			(0.006)	
Second			0.655**	0.34		0.196**	0.10
			(0.007)			(0.006)	
Third			0.693**	0.36		0.330**	0.17
			(0.007)			(0.006)	
Bottom			0.711**	0.37		0.312**	0.16
			(0.007)			(0.006)	

Table 8 (Cont'd): Change in achievement levels by SES quartiles, age categories, and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Top	TIMSS	46,470	0.355**	0.18			
			(0.023)				
Second			0.348**	0.17			
			(0.023)				
Third			0.440**	0.22			
			(0.023)				
Bottom			0.439**	0.22			
			(0.022)				
Top	PISA	28,370	-0.296**	-0.20	24,400	-0.203**	-0.14
			(0.031)			(0.029)	
Second			-0.200**	-0.13		-0.135**	-0.09
			(0.031)			(0.030)	
Third			-0.258**	-0.17		-0.098**	-0.07
			(0.029)			(0.029)	
Bottom			0.067*	0.04		0.218**	0.15
			(0.030)			(0.030)	
Older							
Top	LTT	86,640	0.132**	0.04	65,820	0.046*	0.01
			(0.024)			(0.021)	
Second			0.251**	0.07		-0.033	-0.01
			(0.024)			(0.022)	
Third			0.249**	0.07		-0.064**	-0.02
			(0.024)			(0.022)	
Bottom			0.293**	0.09		0.007	0.00
			(0.024)			(0.023)	
Top	NAEP	68,320	0.133**	0.13	121,800	0.199**	0.10
			(0.017)			(0.015)	
Second			0.050**	0.05		0.043**	0.02
			(0.017)			(0.015)	
Third			-0.013	-0.01		0.100**	0.05
			(0.017)			(0.016)	
Bottom			0.020	0.02		0.068**	0.04
			(0.017)			(0.016)	

Notes & Source: See Table 1. SES has been constructed from constant principal component analysis of parents' education and possession index (see Tables A.4 and A.5). Birth cohorts are: younger age (LTT math 1968-1989 and reading 1970-1989 and NAEP 1980-1993), early adolescence (LTT math 1964-1998 and reading 1966-1998, NAEP 1976-1995, TIMSS 1981-2001 and PISA 1985-2000) and older age (LTT math 1961-1995 and reading 1963-1995 and NAEP math 1988-1998 and reading 1973-1992).

Online Appendix

Consistency of SES effects across different constructs

Parental education

In 12 of the 15 surveys, cohorts of students who have parents with the least amount of education (no high school diploma) outpace those who have the most (college degree or more) educated parents. The differences vary somewhat by the age a student is tested. In math, the median difference is only 0.01sd for those tested at a younger age, 0.10sd for those tested in early adolescence, and 0.05sd for those tested just before completing secondary education. In reading, these numbers are 0.01sd, 0.03sd and -0.05sd. Interestingly, those in the top and bottom categories both make greater progress than do those in the two middle categories (high school and some college) in twelve of the fifteen comparisons (Tables A.11-A.13). This U-shape distribution is evident in both math and reading.

Possession Index

The performance of students in the lowest quartile of the permanent income distribution, as indicated by the number of home possessions, outpaces that of those in the top quartile in nine of seventeen comparisons (Tables A.14-A.16). The top quartile outperforms the bottom quartile in seven comparisons and there is no difference in one survey. The median differential between the top and bottom quartiles across all surveys in both math and in reading is 0.05sd. Nine of the 17 surveys show a U-shaped distribution, with the top and bottom quartiles showing greater progress than the middle two quartiles.

Income (eligibility for free or reduced lunch)

Two surveys—LTT and NAEP—obtain from administrative records a blunt measure of annual income, the eligibility of students for participation in the free or reduced lunch program, which provides subsidized meals to students with incomes below 185 percent of the poverty line. Eligibility in recent years has been gradually extended to all students within some districts if most students are deemed to be eligible. In part for this reason is that the percentage of children eligible for program participation increased between 37 percent in the school year ending in 1999 and 52 percent in the school year ending in 2015 (Chingos, 2016; Greenberg, 2018). Given the broadening of the eligibility definition, the variable is not a constant indicator of annual family income. But as it is an alternative to the possession index, a crude measure of permanent income, we report the rate of progress for those deemed eligible (and ineligible) as reported in twelve LTT and NAEP surveys.

No consistent differentials in the rate of progress are evident for students from higher and lower income households in the twelve surveys for which information is available as to student eligibility for free or reduced lunch. (Table A.17). In six comparisons the ineligible group is making more progress; in four the greater progress is being made by those who are eligible, and in two cases differences are statistically insignificant. Even when significant differences are detected, they are small. Either this indicator of income is too blunt to capture the role of annual income in rates of cohort progress, or the rate of progress is not conditional on annual household income, or the changes in eligibility requirements cloud this measure.

Family structure

In eleven surveys, students report information on the structure of the household in which they live. The surveys categorize household structure in various ways (Table A.6); to facilitate

comparisons within and across surveys we simply dichotomize responses into two broad categories: 1) household where the students say they live with their two parents (not necessarily biological), and 2) all other households. Although family structure is not a precise indicator of social class, two-parent households tend to have higher levels of income (Berger & McLanahan, 2015). The meaning of this indicator may shift with changes in divorce and separation rates.

We find some, but not consistent, signs that the rate of progress is greater for those from households without two parents in the eleven surveys for which this information is available (Table A.18). In seven surveys, the rate of progress for these groups is greater, but in three the opposite is true, and in one survey no significant difference is detected. It could be that the advantage of living in households with two parents has declined over time, or that household formation patterns have changed, or it could be that family structure is not critical for achievement.

References

Berger, L. M., & McLanahan, S. S. (2015). Income, relationship quality, and parenting: Associations with child development in two-parent families. *Journal of Marriage and Family*, *77*(4), 996-1015.

Chingos, M. M. (2016). *No more free lunch for education policymakers and researchers*. Washington, DC: Brookings Institution. Retrieved from: http://www.brookings.edu/research/reports/2016/06/30-no-more-free-lunch-for-education-policy-makers-and-researchers-chingos

Greenberg, E. (2018). *New measures of student poverty: Replacing free and reduced-price lunch status based on household forms with direct certification* [Education policy program policy brief]. Urban Institute.

Table A.1: Selected studies on sociology and economics of education and intelligence.

Study	Datasets	Years	Primary focus	Brief findings
Sociology and economics of education				
Miller (1995)	LTT, other	1971-1990	Ethnic gaps	Some closing of Asian-white and Hispanic-white gaps
Jencks & Phillips (1998)	LTT, other	1971-1996	Black-white gap	Black-white gap closes substantially
Campbell et al. (1996)	LTT, NAEP	1971-1994	Achievement levels	Improvement in math. Not much change in reading
Hedges & Nowell (1995)	NAEP, other	1960-1992	Gender gap	Male advantage in math but not in reading. Gender differences are stable
Hedges & Nowell (1998)	Other	1965-1992	Income, SES, and ethnic gaps	Black-white gap closes. Unclear trend in SES gaps
Grissmer et al. (1994)	Other	1970-1990	Family structure and achievement	Black-white and Hispanic-white gaps close
Phillips, Crouse, & Ralph (1998)	LTT, other	1965-1996	Black-white gap	Gap increases with grades in math, but not in reading
Bloom et al. (2008)	Other	1991-2002	Effect size benchmarks	Growth declines as students move from early grades to later grades
Magnuson, Rosenbaum, & Waldfogel (2008)	LTT	1975-2004	Black-white gap	Black-white gap closes substantially, but little further closing after 1990
Reardon (2011)	Other	1960-2007	90-10 income gap	Large increases in gaps
Hanushek, Peterson, & Woessmann (2012)	NAEP, TIMSS, PIRLS, PISA	1995-2009	Achievement levels	Growth rate is continuous, but modest
Reardon, Valentino, & Shores (2012)	LTT, PIRLS, PISA, other	1971-2009	Literacy patterns	Steeper upward trends in math than reading on the LTT
OECD (2018)	PISA	2000-2015	SES gap	Gaps close modestly in U. S.
Broer, Bai, & Fonseca (2019)	TIMSS	1995-2015	72-25 SES gap	No significant change for gaps in math in U. S.
Chmielewski (2019)	TIMSS, PIRLS, PISA, other	1964-2015	90-10 SES gap	No significant trend in gaps in U. S.

Table A.1 (Cont'd): Selected studies on sociology and economics of education and intelligence.

Study	Datasets	Years	Primary focus	Brief findings
Hanushek et al. (2020)	LTT, NAEP, TIMSS, PISA	1971-2015	75-25 SES gap	Not much change in gaps
Hashim et al. (2020)	NAEP	1990-2015	Income gap	Achievement levels increase. Income gaps close modestly
Shakeel & Peterson (2020)	NAEP	2005-2017	Achievement levels	Larger growth in achievement in charters than at district schools
Matheny et al. (2021)	Other	2009-2018	School district gaps	Gaps across school districts grow based on student SES and ethnic composition
Intelligence studies				
Flynn (1984)	Other	1932-1978	IQ gains	Gains of 3 IQ points per decade
Rodgers & Wänström (2007)	Other	1986-2000	IQ gains	Flynn effect for PIAT-math, but not reading
Ang, Rodgers & Wänström (2010)	Other	1986-2000	IQ gains	Flynn effects for subgroups
Wai & Putallaz (2011)	Other	1981-2000	IQ gains	Flynn effect among high ability students
Pietschnig & Voracek (2015)	Other	1909-2013	IQ gains	Confirm Flynn Effect. Larger increases in fluid reasoning that crystallized knowledge
Rindermann & Thompson (2013)	LTT	1971-2008	IQ gains and ethnic gaps	Larger gains for younger students, and for ethnic minorities
Meisenberg & Woodley (2013)	TIMSS, PISA	1995-2011	IQ gains and cross-country cognitive differences	Higher gains in countries with lower test scores at the beginning of the time period

Note: "Other" datasets include a variety of tests usually administered once or a few times. For details on subjects, grades, exact years for each dataset and samples size, please refer to the original studies.

Table A.2: Test administrations included in analysis, by survey, age, and year of administration.

	LTT (age)		NAEP (grade)		TIMSS (grade)	PIRLS (grade)	PISA (age)
Year	Math	Read.	Math	Read.	Math	Read.	Math/Read.
1971		9,13,17					
1975		9,13,17					
1978	9,13,17						
1980		9,13,17					
1982	9,13,17						
1984		9,13,17					
1986	9,13,17						
1988		9,13,17					
1990	9,13,17	9,13,17	4,8	4,8,12			
1992	9,13,17	9,13,17	4,8	4,8,12			
1994	9,13,17	9,13,17		4,8,12			
1995					4,8		
1996	9,13,17	9,13,17	4,8				
1998				4,8,12			
1999	9,13,17	9,13,17			8		
2000			4,8	4			15
2001						4	
2002				4,8,12			
2003			4,8	4,8	4,8		15
2004	9,13,17	9,13,17					
2005			4,8,12	4,8,12			
2006						4	15*
2007			4,8	4,8	4,8		
2008	9,13,17	9,13,17					
2009			4,8,12	4,8,12			15
2011			4,8	4,8	4,8	4	
2012	9,13,17	9,13,17					15
2013			4,8	4,8			
2015			4,8,12	4,8,12	4,8		15
2016						4	
2017			4,8	4,8			

Note: Table displays math and reading tests available from U.S. Department of Education for each survey. The cells list age/grade of assessment (9,14,15,17;4,8,12). Grade 12 math in NAEP excludes 1990, 1992, 1996 and 2000 as the test format changed in 2005. Student-level LTT math data for 1971 and 1973 could not be obtained. TIMSS was only administered at grade 8 in 1999. We used rescaled scores for the 1995 administration of TIMSS available from https://timssandpirls.bc.edu/timss1995i/Database.html and https://timssandpirls.bc.edu/timss1995i/newscale/layout.txt * PISA test in reading was excluded in 2006 by the administering body. PIRLS 2001 and 2016 waves did not have restricted use versions (information on student ethnicity is missing).
Source: See Table 1.

Table A.3: Ethnic representation in surveys by testing year in math.

Survey	Year	White	Black	Hispanic	Asian	Others
LTT	1978	81	13	5	1	0
	1982	80	13	5	2	0
	1986	73	14	9	1	2
	1990	71	15	10	3	2
	1992	71	15	10	2	1
	1994	70	15	11	3	2
	1996	69	14	12	3	1
	1999	68	14	13	3	2
	2004	63	15	16	4	2
	2008	57	15	21	5	2
	2012	55	14	23	6	2
NAEP	1990	70	15	11	2	2
	1992	69	16	10	3	2
	1996	69	14	12	3	2
	2000	68	14	13	4	2
	2003	61	16	16	4	2
	2005	60	16	17	5	2
	2007	58	16	19	5	2
	2009	57	15	20	5	3
	2011	54	15	22	5	3
	2013	53	15	23	5	3
	2015	51	15	25	6	4
	2017	50	15	25	6	4
TIMSS	1995	66	14	10	4	6
	1999	63	15	12	5	5
	2003	62	15	17	3	3
	2007	55	13	23	4	6
	2011	52	12	25	5	7
	2015	48	12	27	5	8
PISA	2000	59	14	18	3	5
	2003	59	16	17	3	5
	2006	59	13	18	4	6
	2009	56	12	23	5	4
	2012	51	13	25	5	6
	2015	45	13	30	4	6

Note. Table shows weighted ethnic representation in the math test by the year tested. Grades 4 and 8 are combined for TIMSS. Grades 4, 8 and 12 are combined for Main NAEP. Ages 9, 13 and 17 are combined for LTT NAEP. White and Black categories are additionally defined as "non-Hispanic" in TIMSS and PISA. Asians also comprise Pacific Islanders, and we coded it consistently across the surveys. For details see Table A.4.
Source: See Table 1.

Table A.4: Ethnic coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Year	Grade/Age	Coding
LTT* and NAEP	1971-2017	9,13,17 4,8,12	1 White 2 Black 3 Hispanic 4 Asian American/Pacific Islander 5 American Indian/Alaskan Native
TIMSS	1995	4	1 White (not Hispanic) 2 Black (not Hispanic) 3 Hispanic 4 Asian/Pacific Islander 5 American Indian/Alaskan Native 6 Other
		8	1 White 2 Black 3 Hispanic 4 Asian/Pacific Islander 5 Native American 6 Other
	1999	8	1 White 2 Black 3 Hispanic 4 Asian/Pacific Islander 5 Native American 6 Other 8 Not admin. 9 missing
	2003	4,8	0 Missing 1 White 2 Black 3 Hispanic 4 Asian 5 American Indian/Native Hawaiian/Pacific Islander
	2007, 2011, 2015	4,8	1 White, not Hispanic 2 Black, not Hispanic 3 Hispanic 4 Asian 5 Native American 6 Pacific Islander 7 2 or more races 9 omitted
PIRLS	2006	4	1 Hispanic 2 White, not Hispanic 3 Black, not Hispanic 4 Asian, not Hispanic 5 American Indian/Alaskan, not Hispanic 6 Multiple races, not Hispanic 9 omitted
	2011		1 White, not Hispanic 2 Black, not Hispanic 3 Hispanic 4 Asian 5 Native American 6 Pacific islander 7 2 or more races 9 omitted or invalid
PISA	2000	15	1 Hispanic 2 White 3 Black 4 Asian 5 American Indian 6 Hawaiian 7 Multirace 9 Missing
	2003		0 Missing 1 White 2 Black 3 Hispanic 4 Asian 5 Other 6 Multirace
	2006		1 Hispanic 2 White, not Hispanic 3 Black, not Hispanic 4 Asian, not Hispanic 5 Multiracial, not Hispanic 6 American Indian, not Hispanic 7 Pacific Islander, not Hispanic 9 Missing
	2009		1 White 2 Black 3 Hispanic 4 Asian 5 American Indian/Alaska Native 6 Native Hawaiian/Other Pacific Islander 7 More than one race 8 N/A 9 Miss
	2012		1 White 2 Black or African American 3 Hispanic 4 Asian 5 American Indian or Alaska Native 6 Native Hawaiian or Other Pacific Islander 7 More than one race 97 N/A 99 Miss

Table A.4 (Cont'd): Ethnic coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Year	Grade/Age	Coding
PISA	2015	15	1 White, not Hispanic 2 Black or African American 3 Hispanic or Latino 4 Asian 5 American Indian or Alaska Native 6 Native Hawaiian or Other Pacific 7 Multi-Racial 99 No Response

Note. Table displays codes and labels for ethnicity across the surveys. Some datasets in NAEP do no mention "Pacific Islander" and "Alaskan Native" in the variable label. * For 1971 LTT reading does not have Asian and Other categories. Restricted-use data on ethnic identification for PIRLS 2001 and 2016 was not available from NCES.
Source: See Table 1.

Table A.5: Parent's education coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Years	Notes
LTT	1971-2012	At age 9, parent's education is only reported for 1971-1999. Parent's education is coded as 1 Did not finish High School 2 Graduated High School 3 Post High School 4 Graduated college. In 1971 and 1975, the parent's education variable PARED4 was constructed from a variable on parent's highest education PEDH (values 0,1,2,3,5,9). The tabulations of the two variables mismatch for 1971 and 1975 (graduated college category in PARED4 comprises of both 0 and 9 values in PEDH). Documentation is lacking to resolve these discrepancies. Thus, PARED4 is poorly constructed. Hence, we excluded data on parent's education for 1971 and 1975.
NAEP	1990-2017	For grade 4, parent's education is only reported for 1990-1998, 2000 Reading and 2003. Parent's education is coded as 1 Did not finish High School 2 Graduated High School 3 Post High School 4 Graduated college.
TIMSS	1995-2015	Parent's education is not reported for grade 4. Parent's education is coded as 1 No more than primary schooling 2 Finished lower secondary schooling 3 Finished upper secondary schooling 4 Finish post-secondary, vocational/technical education but no university 5 Finished university/equivalent or higher. Parent's education can be recoded to compare to Main NAEP and LTT NAEP. For parent's education the coding would be (2=1) (3=2) (4=3) (5=4).
PISA	2000-2015	Parent's education is coded as 0 None 1 ISCED 1 i.e. elementary school 2 ISCED 2 i.e. middle/junior high school 3 ISCED 3B, C i.e. high school equivalent/GED 4 ISCED 3A, ISCED 4 i.e. high school diploma/ vocational or other certificate/ diploma of high school 5 ISCED 5B i.e. an associate degree 6 ISCED 5A, 6 i.e. bachelor, master, doctorate, professional degree, law, or medicine. Parent's education can be recoded to compare to Main NAEP and LTT NAEP. For parent's education the coding would be (0 2 =1) (3 4 =2) (5=3) (6=4).

Source: See Table 1.

Table A.6: Possession index coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Years	Items	Total items	Notes
LTT	1971-2012	Books	4	Books is binary for 1971-1999. The binary measure has coding 0 “<=25 books” 1 “>25 books.” Books has four categories for 2004-2012. The four categories are 1 "0-10 i.e. few” 2 "11-25 i.e. one shelf” 3 "26-100 i.e. one bookcase 4 ">100 i.e. > one bookcase.” To maintain a consistent measurement, we recoded books as binary across all the years within LTT.
	1971-2012	Newspaper, magazines, encyclopedia		Encyclopedia is missing in 1984.
NAEP	1990-2017	Books	4	Books is binary for 1990-2000. The binary measure has coding 0 “<=25 books” 1 “>25 books.” Books has four categories after the year 2000. The four categories are 1 "0-10 i.e. few” 2 "11-25 i.e. one shelf” 3 "26-100 i.e. one bookcase 4 ">100 i.e. > one bookcase.” To maintain a consistent measurement, we recoded books as binary across all the years within NAEP.
	1990-2010	Newspaper, magazines, encyclopedia		Magazine and encyclopedia exist in 2011
	2005-2015	Computer	2	This variable was only used for grade 12 Math.
TIMSS	1995-2015	Books	7	Books has five categories. The five categories are 1 "0-10" 2 "11-25" 3 "26-100" 4 "101-200" 5 ">200."
	1995-2015	Study desk, computer		Using own computer as proxy for computer in 2015
PIRLS	2001-2016	Books	8	Books has five categories. The five categories are 1 "0-10" 2 "11-25" 3 "26-100" 4 "101-200" 5 ">200."
	2001-2016	Study desk, computer, own books		

Table A.6 (Cont'd): Possession index coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Years	Items	Total items	Notes
PISA	2000-2015	Books	25	Books has six categories. The six categories are 1 "0-10" 2 "11-25" 3 "26-100" 4 "101-200" 5 "251-500" 6"More than 500." PISA 2000 has seven categories (1 "none" 2 "0-10" 3"11-50" 4 "51-100" 5 "101-200" 6 "251-500" 7"More than 500"). We assumed that the weighted proportional distribution of #books remains same for 2000 and 2003, and recoded books as (2=1) (3=2) (4=3) (5=4) (6=5) (7=6).
	2000-2015	Study desk, own room, study place, computer, educational software, internet, classic literature, poetry books, art works (e.g. paintings), textbooks, dictionary, TV, car		

Note: All items in LTT and NAEP are dichotomous. In TIMSS and PIRLS, except for books at home, all other items are dichotomous. In PISA, books are categorical, and the following items are polytomous and coded 0-3: computer, TV, and car. All remaining items are dichotomous.
Source: See Table 1.

Table A.7: Family structure coding in the test administrations included in analysis, by survey, age, and year of administration.

Survey	Year	Grade/Age	Coding
LTT	1986, 1990, 1992, 1994, 1996, 1999	9,13, 17	1 2 parents at home 2 1 parent at home 3 Neither 8 Omitted 9 Not reached
NAEP	1990,1992, 1994	4,8,12	1 2 Parents at home 2 1 Parent at home 3 Neither Parent Home 8 Omitted
	1996*		Does (mother/father) or (stepmother/stepfather) live at home with you? 0 Multiple Response 1 Yes 2 No 8 Omitted
	2013*	4, 8	Lives in home: (mother/ stepmother/ father/ stepfather)? 1 Yes 8 Omitted
	2015*	4, 8, 12	
	2017*	4, 8	Lives in home: (mother/ stepmother/ father/ stepfather)? 1 Yes 2 No 8 Omitted
TIMSS	1995*	4,8	Student lives with (mother/ stepmother/ father/ stepfather)? 1 yes 2 no 8 not admin. 9 missing
	1999*	8	
PISA	2000, 2003	15	1 Single parent family 2 Nuclear family 3 Mixed family 4 Other 9 Missing
	2009		1 Single parent (natural or otherwise) 2 Two parents (natural or otherwise) 3 Other 7 N/A 9 Missing
	2012		1 Single parent (natural or otherwise) 2 Two parents (natural or otherwise) 3 Other 9 Missing

Note: Table displays the available data for family structure. Data on family structure was only collected for Math in LTT. For the years marked with an asterisk (*), family structure variable was coded from questions about who lives in the home.
Source: See Table 1.

Table A.8: Medians of change/decade in achievement levels at younger age and early adolescence in cross-country surveys.

(1)	(2)	(3)	(4)	(5)	(6)
Survey	Sample	Math	Reading	Math	Reading
		Younger age (grade 4)		Early adolescence (grade 8/age 15)	
TIMSS	Overall	0.24		0.09	
	OECD	0.21		0.08	
	Non-OECD	0.28		0.09	
N		340,130		386,650	
PIRLS	Overall		0.05		
	OECD		0.02		
	Non-OECD		0.19		
N			307,980		
PISA	Overall			-0.04	-0.01
	OECD			-0.05	-0.01
	Non-OECD			0.07	0.04
N				1,728,980	1,767,990

Note: Table shows medians of change/decade in achievement levels for 10 regions in TIMSS, 32 regions in PISA and 16 regions in PIRLS. Normalized achievement is measured in standard deviations (s.d.). The s.d. is the difference between the year test was administered and the starting year for a specific set of countries that were consistently tested for all administrations of a test between the available years. The changes have been estimated from a quadratic fit. See Tables A.9-A.11 for cross-country comparisons.
For TIMSS, 10 countries were tested both at grades 4 and 8 between 1995-2019 and these countries also has test data for the initial and final years (https://nces.ed.gov/timss/participation.asp). For PISA, 32 regions were tested in all years between 2001-2018 (https://nces.ed.gov/surveys/pisa/countries.asp). For PIRLS, 16 regions were tested for all years between 2001-2016 (https://nces.ed.gov/surveys/pirls/countries.asp). We used rescaled scores for the 1995 administration of TIMSS available from https://timssandpirls.bc.edu/timss1995i/Database.html and https://timssandpirls.bc.edu/timss1995i/newscale/layout.txt PISA test in reading was excluded in 2006 in U. S. by the administering body.
OECD represents regions that are affiliated with the Organisation for Economic Co-operation and Development.
SOURCE: The International Association for the Evaluation of Educational Achievement (IEA), TIMSS 1995-2019, PISA 2000-2018 and PIRLS 2001-2016.

Table A.9: Change in achievement levels at younger age and early adolescence in cross-country TIMSS survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Grade 4			Grade 8		
Country	Change	s.e.	Change/dc	Change	s.e.	Change/dc
Australia	0.296**	(0.012)	0.12	0.029**	(0.011)	0.01
Hong Kong-China*	0.709**	(0.012)	0.30	0.189**	(0.014)	0.08
Hungary	0.321**	(0.014)	0.13	-0.066**	(0.014)	-0.03
Iran*	0.788**	(0.014)	0.33	0.398**	(0.013)	0.17
Japan	0.517**	(0.011)	0.22	0.213**	(0.013)	0.09
New Zealand	0.408**	(0.014)	0.17	-0.010	(0.013)	0.00
Norway	1.164**	(0.014)	0.49	0.350**	(0.012)	0.15
Singapore*	0.616**	(0.011)	0.26	0.232**	(0.012)	0.10
United Kingdom	0.868**	(0.015)	0.36	0.322**	(0.015)	0.13
United States	0.476**	(0.009)	0.20	0.323**	(0.010)	0.13

Notes & Source: See Table A.8. Table displays cross-country changes in achievement levels in TIMSS between 1995-2019. The changes have been estimated from a quadratic fit. Change per decade (dc) in achievement levels is also displayed. Standard errors are in parenthesis. *p <0.05, **p<0.01.* represents Non-OECD region. Hong Kong is not an independent country and therefore is not discussed in text.

Table A.10: Change in achievement levels at younger age in cross-country PIRLS survey.

(1)	(2)	(3)	(4)
Country	Change	s.e.	Change/dc
Bulgaria*	-0.038*	(0.019)	-0.03
France	-0.144**	(0.015)	-0.10
Germany	-0.043**	(0.013)	-0.03
Hong Kong-China*	0.425**	(0.014)	0.28
Hungary	0.067**	(0.015)	0.04
Iran*	0.278**	(0.018)	0.19
Italy	0.054**	(0.016)	0.04
Lithuania*	0.038**	(0.015)	0.03
Netherlands	-0.087**	(0.013)	-0.06
New Zealand	-0.062**	(0.019)	-0.04
Russian Federation*	0.546**	(0.014)	0.36
Singapore*	0.501**	(0.015)	0.33
Slovakia*	0.187**	(0.015)	0.12
Slovenia*	0.426**	(0.015)	0.28
United Kingdom	0.094**	(0.019)	0.06
United States	0.116**	(0.014)	0.08

Notes & Source: See Table A.8. Table displays cross-country changes in achievement levels in PIRLS between 2001-2016. The changes have been estimated from a quadratic fit. Change per decade (dc) in achievement levels is also displayed. Standard errors are in parenthesis. *p <0.05, **p<0.01.* represents Non-OECD region. Hong Kong is not an independent country and therefore is not discussed in text.

Table A.11: Change in achievement levels at early adolescence in cross-country PISA survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Math			Reading		
Country	Change	s.e.	Change/dc	Change	s.e.	Change/dc
Australia	-0.375**	(0.008)	-0.21	-0.249**	(0.009)	-0.14
Belgium	-0.150**	(0.011)	-0.08	-0.120**	(0.012)	-0.07
Brazil*	0.380**	(0.007)	0.21	0.155**	(0.009)	0.09
Canada	-0.186**	(0.006)	-0.10	-0.105**	(0.006)	-0.06
Czech Republic	-0.096**	(0.012)	-0.05	-0.011	(0.013)	-0.01
Denmark	-0.077**	(0.011)	-0.04	0.056**	(0.012)	0.03
Finland	-0.319**	(0.011)	-0.18	-0.278**	(0.012)	-0.15
France	-0.190**	(0.013)	-0.11	-0.029	(0.015)	-0.02
Germany	0.090**	(0.013)	0.05	0.162**	(0.015)	0.09
Greece	0.034**	(0.013)	0.02	-0.105**	(0.014)	-0.06
Hong Kong-China*	-0.025	(0.014)	-0.01	0.060**	(0.014)	0.03
Hungary	-0.103**	(0.013)	-0.06	-0.075**	(0.014)	-0.04
Iceland	-0.224**	(0.015)	-0.12	-0.241**	(0.018)	-0.13
Indonesia*	0.122**	(0.009)	0.07	0.064**	(0.008)	0.04
Ireland	-0.012	(0.012)	-0.01	-0.022	(0.014)	-0.01
Italy	0.288**	(0.007)	0.16	0.004	(0.008)	0.00
Korea	-0.173**	(0.013)	-0.10	-0.171**	(0.013)	-0.10
Latvia	0.158**	(0.013)	0.09	0.156**	(0.015)	0.09
Luxembourg	0.184**	(0.014)	0.10	0.221**	(0.016)	0.12
Mexico	0.216**	(0.005)	0.12	0.098**	(0.006)	0.05
Netherlands	-0.338**	(0.014)	-0.19	-0.333**	(0.015)	-0.19
New Zealand	-0.364**	(0.014)	-0.20	-0.201**	(0.016)	-0.11
Norway	0.016	(0.013)	0.01	0.061**	(0.016)	0.03
Poland	0.337**	(0.014)	0.19	0.262**	(0.015)	0.15
Portugal	0.334**	(0.012)	0.19	0.247**	(0.013)	0.14
Russian Federation*	0.147**	(0.012)	0.08	0.367**	(0.013)	0.20
Spain	0.044**	(0.006)	0.02	0.283**	(0.012)	0.16
Sweden	-0.146**	(0.014)	-0.08	-0.161**	(0.015)	-0.09
Switzerland	-0.096**	(0.010)	-0.05	-0.086**	(0.011)	-0.05
Thailand*	-0.059**	(0.010)	-0.03	-0.221**	(0.010)	-0.12
United Kingdom	-0.212**	(0.008)	-0.12	-0.147**	(0.010)	-0.08
United States	-0.106**	(0.013)	-0.06	0.020	(0.015)	0.01

Notes & Source: See Table A.8. Table displays cross-country changes in achievement levels in PISA between 2000-2018. PISA test in reading was excluded in 2006 by the administering body. The changes have been estimated from a quadratic fit. Change per decade (dc) in achievement levels is also displayed. Standard errors are in parenthesis. *p <0.05, **p<0.01. * represents Non-OECD country. Hong Kong is not an independent country and therefore is not discussed in text.

Table A.12: Change in achievement levels at younger age by gender and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Gender	N	Change	Change/dc	N	Change	Change/dc
LTT	Male	44,340	0.822**	0.24	68,780	0.298**	0.07
			(0.017)			(0.015)	
	Female	44,830	0.745**	0.22	68,970	0.147**	0.04
			(0.016)			(0.015)	
NAEP	Male	740,100	1.054**	0.39	812,600	0.273**	0.10
			(0.004)			(0.003)	
	Female	715,530	1.050**	0.39	793,430	0.170**	0.06
			(0.003)			(0.003)	
TIMSS	Male	25,490	0.640**	0.32			
			(0.016)				
	Female	25,950	0.581**	0.29			
			(0.015)				
PIRLS	Male				12,980	0.176**	0.12
						(0.022)	
	Female				13,130	0.086**	0.06
						(0.021)	

Notes & Source: See Table 1.

Table A.13: Change in achievement levels at early adolescence by gender and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Gender	N	Change	Change/dc	N	Change	Change/dc
LTT	Male	50,300	0.527**	0.16	70,930	0.152**	0.04
			(0.017)			(0.016)	
	Female	50,900	0.457**	0.13	71,140	0.105**	0.03
			(0.016)			(0.016)	
NAEP	Male	658,950	0.730**	0.27	716,270	0.391**	0.14
			(0.004)			(0.003)	
	Female	643,000	0.701**	0.26	707,290	0.247**	0.09
			(0.004)			(0.003)	
TIMSS	Male	28,110	0.359**	0.18			
			(0.016)				
	Female	28,920	0.390**	0.20			
			(0.015)				
PISA	Male	14,680	-0.140**	-0.09	12,640	0.003	0.00
			(0.024)			(0.023)	
	Female	14,450	-0.152**	-0.10	12,590	-0.059**	-0.04
			(0.022)			(0.021)	

Notes & Source: See Table 1.

Table A.14: Change in achievement levels at older age by gender and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Math			Reading		
Survey	Gender	N	Change	Change/dc	N	Change	Change/dc
LTT	Male	45,420	0.180**	0.05	68,380	0.003	0.00
			(0.018)			(0.015)	
	Female	47,050	0.250**	0.07	68,980	0.031*	0.01
			(0.018)			(0.015)	
NAEP	Male	35,590	0.052**	0.05	71,860	0.081**	0.03
			(0.014)			(0.012)	
	Female	35,900	0.039**	0.04	73,300	0.053**	0.02
			(0.013)			(0.011)	

Notes & Source: See Table 1.

Table A.15: Medians of change/decade in achievement levels in math and reading by SES quartile for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Ethnicity	Base	SES quart	Young.	Diff.	E.A.	Diff.	Older	Diff.	All ages	Diff.
Math										
Asian	Top	Top	0.24		0.21		0.09		0.14	
		Second	0.42	0.18	0.26	0.06	0.05	-0.05	0.24	0.10
		Third	0.48	0.25	0.30	0.09	0.15	0.06	0.24	0.10
		Bottom	0.42	0.19	0.28	0.08	0.19	0.09	0.28	0.15
Black	Top	Top	0.48		0.27		0.19		0.29	
		Second	0.55	0.07	0.26	-0.01	0.13	-0.06	0.26	-0.03
		Third	0.56	0.08	0.28	0.02	0.08	-0.12	0.28	-0.01
		Bottom	0.59	0.11	0.24	-0.03	0.11	-0.09	0.24	-0.05
Hispanic	Top	Top	0.42		0.26		0.18		0.21	
		Second	0.40	-0.02	0.28	0.02	0.10	-0.08	0.17	-0.04
		Third	0.48	0.06	0.26	0.00	0.18	0.00	0.21	0.00
		Bottom	0.51	0.10	0.29	0.03	0.19	0.01	0.23	0.02
White	Top	Top	0.45		0.15		0.11		0.17	
		Second	0.46	0.01	0.18	0.03	0.11	-0.01	0.18	0.01
		Third	0.49	0.04	0.20	0.05	0.06	-0.05	0.20	0.03
		Bottom	0.50	0.04	0.20	0.05	0.07	-0.04	0.20	0.03
Reading										
Asian	Top	Top	0.03		-0.02		0.05		0.03	
		Second	-0.04	-0.07	-0.01	0.01	0.16	0.11	0.05	0.01
		Third	0.15	0.12	0.05	0.06	0.13	0.08	0.08	0.05
		Bottom	0.09	0.06	0.10	0.11	0.17	0.12	0.16	0.13
Black	Top	Top	0.01		0.07		0.08		0.07	
		Second	0.05	0.04	0.09	0.02	0.13	0.05	0.10	0.02
		Third	0.07	0.06	0.13	0.05	0.09	0.01	0.13	0.05
		Bottom	0.13	0.12	0.13	0.06	0.09	0.01	0.12	0.05
Hispanic	Top	Top	0.01		0.11		0.05		0.05	
		Second	-0.02	-0.03	0.14	0.03	0.04	0.00	0.05	0.00
		Third	0.07	0.06	0.16	0.05	0.07	0.02	0.09	0.03
		Bottom	0.18	0.17	0.28	0.17	0.16	0.11	0.21	0.16
White	Top	Top	-0.05		0.05		0.08		0.03	
		Second	0.00	0.04	0.04	-0.01	0.03	-0.05	0.03	0.00
		Third	-0.01	0.04	0.04	-0.01	0.02	-0.05	0.02	-0.01
		Bottom	0.10	0.15	0.05	0.00	-0.02	-0.10	0.05	0.01

Notes & Source: See Table 1. Table displays medians of change/decade in achievement levels by SES quartile for each ethnic group displayed in Tables A.16 - A.18. E. A. = Early Adolescence. Birth years differ by age categories, depending on the availability of data (see Table 8 for details). Differences between the base category and other categories are also displayed.

Table A.16: Change in achievement levels at younger age by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	LTT	800	0.024	0.01	540	0.082	0.04
				(0.204)			(0.201)	
	Second			0.541**	0.26		-0.244	-0.13
				(0.183)			(0.173)	
	Third			0.254	0.12		0.363	0.19
				(0.177)			(0.189)	
	Bottom			0.642**	0.31		-0.137	-0.07
				(0.205)			(0.227)	
Black	Top		7,690	0.656**	0.31	4,700	-0.129	-0.07
				(0.069)			(0.083)	
	Second			0.725**	0.35		0.020	0.01
				(0.072)			(0.086)	
	Third			0.812**	0.39		-0.007	0.00
				(0.073)			(0.088)	
	Bottom			0.817**	0.39		0.052	0.03
				(0.072)			(0.092)	
Hispanic	Top		5,510	0.413**	0.20	3,630	0.159	0.08
				(0.076)			(0.090)	
	Second			0.327**	0.16		-0.018	-0.01
				(0.079)			(0.086)	
	Third			0.389**	0.19		0.129	0.07
				(0.078)			(0.091)	
	Bottom			0.337**	0.16		0.262**	0.14
				(0.076)			(0.094)	
White	Top		31,520	0.571**	0.27	21,470	-0.100*	-0.05
				(0.036)			(0.041)	
	Second			0.658**	0.31		0.055	0.03
				(0.036)			(0.042)	
	Third			0.650**	0.31		-0.082	-0.04
				(0.037)			(0.043)	
	Bottom			0.615**	0.29		0.089	0.05
				(0.039)			(0.045)	

Table A.16 (Cont'd): Change in achievement levels at younger age by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	NAEP	5,620	0.598**	0.46	6,030	0.021	0.02
				(0.063)			(0.058)	
	Second			0.747**	0.57		0.059	0.05
				(0.065)			(0.057)	
	Third			1.098**	0.84		0.131*	0.10
				(0.065)			(0.065)	
	Bottom			0.700**	0.54		0.336**	0.26
				(0.059)			(0.060)	
Black	Top		28,180	0.849**	0.65	30,530	0.122**	0.09
				(0.030)			(0.028)	
	Second			0.984**	0.76		0.124**	0.10
				(0.027)			(0.028)	
	Third			0.948**	0.73		0.182**	0.14
				(0.030)			(0.029)	
	Bottom			1.029**	0.79		0.307**	0.24
				(0.029)			(0.029)	
Hispanic	Top		16,130	0.826**	0.64	18,210	-0.085*	-0.07
				(0.037)			(0.039)	
	Second			0.825**	0.63		-0.031	-0.02
				(0.036)			(0.039)	
	Third			0.996**	0.77		0.083*	0.06
				(0.036)			(0.039)	
	Bottom			1.124**	0.86		0.279**	0.21
				(0.036)			(0.038)	
White	Top		93,670	0.826**	0.64	101,990	-0.049**	-0.04
				(0.016)			(0.014)	
	Second			0.799**	0.61		-0.041**	-0.03
				(0.016)			(0.015)	
	Third			0.875**	0.67		0.030*	0.02
				(0.016)			(0.015)	
	Bottom			0.913**	0.70		0.211**	0.16
				(0.016)			(0.015)	

Notes & Source: See Table 1 and Table 8. Also see Table 8 for birth cohorts.

Table A.17: Change in achievement levels at early adolescence by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	LTT	2,630	0.323**	0.10	2,000	0.433**	0.14
				(0.081)			(0.126)	
	Second			1.043**	0.31		0.371**	0.12
				(0.089)			(0.127)	
	Third			1.129**	0.33		0.252*	0.08
				(0.094)			(0.122)	
	Bottom			0.878**	0.26		0.312*	0.10
				(0.096)			(0.129)	
Black	Top		13,600	0.672**	0.20	9,310	0.228**	0.07
				(0.053)			(0.063)	
	Second			0.739**	0.22		0.302**	0.09
				(0.053)			(0.064)	
	Third			0.767**	0.23		0.403**	0.13
				(0.053)			(0.065)	
	Bottom			0.730**	0.21		0.417**	0.13
				(0.053)			(0.067)	
Hispanic	Top		10,010	0.681**	0.20	7,550	0.074	0.02
				(0.048)			(0.062)	
	Second			0.637**	0.19		0.265**	0.08
				(0.048)			(0.061)	
	Third			0.793**	0.23		0.369**	0.12
				(0.047)			(0.064)	
	Bottom			0.800**	0.24		0.405**	0.13
				(0.046)			(0.063)	
White	Top		62,630	0.441**	0.13	44,120	0.158**	0.05
				(0.027)			(0.031)	
	Second			0.622**	0.18		0.125**	0.04
				(0.027)			(0.031)	
	Third			0.641**	0.19		0.133**	0.04
				(0.028)			(0.031)	
	Bottom			0.608**	0.18		0.156**	0.05
				(0.029)			(0.033)	

Table A.17 (Cont'd): Change in achievement levels at early adolescence by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	NAEP	25,220	0.644**	0.34	30,330	-0.030	-0.02
				(0.036)			(0.032)	
	Second			0.771**	0.41		-0.014	-0.01
				(0.039)			(0.031)	
	Third			0.493**	0.26		0.093**	0.05
				(0.037)			(0.031)	
	Bottom			0.590**	0.31		0.181**	0.10
				(0.034)			(0.029)	
Black	Top		98,830	0.804**	0.42	117,810	0.251**	0.13
				(0.017)			(0.014)	
	Second			0.971**	0.51		0.338**	0.18
				(0.017)			(0.015)	
	Third			0.832**	0.44		0.353**	0.19
				(0.017)			(0.014)	
	Bottom			0.944**	0.50		0.365**	0.19
				(0.016)			(0.014)	
Hispanic	Top		70,950	0.712**	0.37	80,080	0.359**	0.19
				(0.020)			(0.018)	
	Second			0.773**	0.41		0.363**	0.19
				(0.020)			(0.018)	
	Third			0.789**	0.42		0.340**	0.18
				(0.019)			(0.018)	
	Bottom			0.899**	0.47		0.523**	0.28
				(0.019)			(0.018)	
White	Top		375,350	0.752**	0.40	454,720	0.324**	0.17
				(0.008)			(0.007)	
	Second			0.663**	0.35		0.239**	0.13
				(0.008)			(0.007)	
	Third			0.705**	0.37		0.314**	0.17
				(0.008)			(0.007)	
	Bottom			0.747**	0.39		0.398**	0.21
				(0.008)			(0.007)	

Table A.17 (Cont'd): Change in achievement levels at early adolescence by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	TIMSS	1,990	0.644**	0.32			
				(0.112)				
	Second			0.442**	0.22			
				(0.115)				
	Third			0.784**	0.39			
				(0.108)				
	Bottom			0.706**	0.35			
				(0.113)				
Black	Top		6,560	0.674**	0.34			
				(0.065)				
	Second			0.613**	0.31			
				(0.062)				
	Third			0.687**	0.34			
				(0.060)				
	Bottom			0.524**	0.26			
				(0.058)				
Hispanic	Top		8,960	0.650**	0.33			
				(0.053)				
	Second			0.745**	0.37			
				(0.051)				
	Third			0.574**	0.29			
				(0.048)				
	Bottom			0.692**	0.35			
				(0.047)				
White	Top		26,090	0.355**	0.18			
				(0.029)				
	Second			0.358**	0.18			
				(0.029)				
	Third			0.425**	0.21			
				(0.028)				
	Bottom			0.454**	0.23			
				(0.028)				

Table A.17 (Cont'd): Change in achievement levels at early adolescence by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	PISA	1,160	0.139	0.09	1,010	-0.103	-0.07
				(0.151)			(0.139)	
	Second			-0.588**	-0.39		-0.283	-0.19
				(0.173)			(0.159)	
	Third			-0.065	-0.04		-0.248	-0.17
				(0.168)			(0.158)	
	Bottom			0.196	0.13		0.288	0.19
				(0.168)			(0.170)	
Black	Top		3,940	0.027	0.02	3,490	-0.017	-0.01
				(0.075)			(0.078)	
	Second			-0.036	-0.02		0.017	0.01
				(0.073)			(0.077)	
	Third			0.029	0.02		0.113	0.08
				(0.069)			(0.070)	
	Bottom			0.047	0.03		0.093	0.06
				(0.067)			(0.072)	
Hispanic	Top		6,310	0.056	0.04	5,620	0.158*	0.11
				(0.069)			(0.064)	
	Second			0.074	0.05		0.210**	0.14
				(0.062)			(0.062)	
	Third			0.137*	0.09		0.236**	0.16
				(0.062)			(0.062)	
	Bottom			0.236**	0.16		0.533**	0.36
				(0.061)			(0.061)	
White	Top		15,200	-0.302**	-0.20	12,770	-0.207**	-0.14
				(0.040)			(0.039)	
	Second			-0.200**	-0.13		-0.105**	-0.07
				(0.040)			(0.039)	
	Third			-0.234**	-0.16		-0.094*	-0.06
				(0.038)			(0.038)	
	Bottom			-0.158**	-0.11		-0.035	-0.02
				(0.040)			(0.040)	

Notes & Source: See Table 1 and Table 8. Also see Table 8 for birth cohorts.

Table A.18: Change in achievement levels at older age by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	LTT	2,560	0.036	0.01	2,290	0.105	0.03
				(0.095)			(0.107)	
	Second			-0.056	-0.01		0.223*	0.07
				(0.097)			(0.110)	
	Third			0.288**	0.08		0.156	0.05
				(0.095)			(0.105)	
	Bottom			0.578**	0.17		0.508**	0.16
				(0.097)			(0.114)	
Black	Top		12,130	0.408**	0.12	9,000	0.315**	0.10
				(0.059)			(0.054)	
	Second			0.591**	0.17		0.433**	0.14
				(0.059)			(0.056)	
	Third			0.592**	0.17		0.409**	0.13
				(0.058)			(0.059)	
	Bottom			0.553**	0.16		0.396**	0.12
				(0.056)			(0.058)	
Hispanic	Top		8,320	0.473**	0.14	7,110	0.162**	0.05
				(0.053)			(0.056)	
	Second			0.467**	0.14		0.115*	0.04
				(0.052)			(0.058)	
	Third			0.586**	0.17		0.273**	0.09
				(0.051)			(0.060)	
	Bottom			0.730**	0.21		0.329**	0.10
				(0.051)			(0.059)	
White	Top		62,750	0.188**	0.06	46,520	0.108**	0.03
				(0.029)			(0.027)	
	Second			0.326**	0.10		0.002	0.00
				(0.030)			(0.027)	
	Third			0.366**	0.11		0.018	0.01
				(0.030)			(0.027)	
	Bottom			0.309**	0.09		-0.077**	-0.02
				(0.031)			(0.029)	

Table A.18 (Cont'd): Change in achievement levels at older age by SES quartile and survey for each ethnic group.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Math			Read.		
Ethnicity	Quartile	Survey	N	Change	Change/dc	N	Change	Change/dc
Asian	Top	NAEP	3,290	0.178*	0.18	5,250	0.130	0.07
				(0.087)			(0.074)	
	Second			0.106	0.11		0.465**	0.24
				(0.079)			(0.073)	
	Third			0.216*	0.22		0.410**	0.22
				(0.083)			(0.077)	
	Bottom			0.201*	0.20		0.360**	0.19
				(0.089)			(0.081)	
Black	Top		9,540	0.267**	0.27	17,570	0.128**	0.07
				(0.046)			(0.039)	
	Second			0.086*	0.09		0.252**	0.13
				(0.044)			(0.038)	
	Third			-0.024	-0.02		0.112**	0.06
				(0.042)			(0.038)	
	Bottom			0.052	0.05		0.099*	0.05
				(0.043)			(0.039)	
Hispanic	Top		10,180	0.219**	0.22	15,180	0.080	0.04
				(0.044)			(0.042)	
	Second			0.061	0.06		0.099*	0.05
				(0.045)			(0.042)	
	Third			0.178**	0.18		0.107*	0.06
				(0.045)			(0.043)	
	Bottom			0.161**	0.16		0.410**	0.22
				(0.042)			(0.045)	
White	Top		44,170	0.170**	0.17	81,710	0.225**	0.12
				(0.021)			(0.018)	
	Second			0.115**	0.12		0.114**	0.06
				(0.020)			(0.018)	
	Third			0.019	0.02		0.083**	0.04
				(0.020)			(0.019)	
	Bottom			0.058**	0.06		-0.030	-0.02
				(0.020)			(0.020)	

Notes & Source: See Table 1 and Table 8. Also see Table 8 for birth cohorts.

Table A.19: Medians of change/decade in achievement levels in math and reading for different SES constructs by subgroups and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Base	Subgroup	Young	Diff.	E. A.	Diff.	Older	Diff.	All age	Diff.
Math									
Pared.: Grad. coll.	Grad. coll.	0.44		0.11		0.05		0.11	
	Post h.s.	0.38	-0.06	0.08	-0.03	0.01	-0.04	0.08	-0.03
	Grad. h.s.	0.35	-0.09	0.12	0.01	0.02	-0.03	0.09	-0.02
	< h.s.	0.45	0.01	0.21	0.10	0.10	0.05	0.21	0.10
Poss. quart: Top	Top	0.31		0.18		0.14		0.19	
	Second	0.28	-0.03	0.17	-0.01	0.10	-0.04	0.17	-0.02
	Third	0.26	-0.05	0.15	-0.03	-0.11	-0.25	0.15	-0.04
	Bottom	0.36	0.05	0.24	0.06	0.08	-0.06	0.24	0.05
FRL: No	No	0.32		0.27		0.17		0.26	
	Yes	0.31	-0.01	0.27	0.00	0.15	-0.02	0.25	-0.01
Two parents: Yes	Yes	0.29		0.20		0.12		0.20	
	No	0.32	0.03	0.13	-0.07	0.17	0.05	0.22	0.02
Reading									
Pared.: Grad. coll.	Grad. coll.	0.05		0.10		0.08		0.10	
	Post h.s.	-0.08	-0.13	0.01	-0.09	-0.05	-0.13	-0.02	-0.12
	Grad. h.s.	0.03	-0.02	0.08	-0.02	-0.04	-0.12	-0.02	-0.12
	< h.s.	0.06	0.01	0.13	0.03	0.03	-0.05	0.06	-0.04
Poss. quart: Top	Top	0.05		0.10		0.08		0.06	
	Second	0.09	0.04	0.07	-0.03	0.03	-0.05	0.05	-0.01
	Third	0.05	0.00	0.05	-0.05	-0.03	-0.11	0.02	-0.04
	Bottom	0.11	0.06	0.09	-0.01	-0.01	-0.09	0.11	0.05
FRL: No	No	0.18		0.24		0.15		0.18	
	Yes	0.19	0.01	0.17	-0.07	0.16	0.01	0.19	0.01
Two parents: Yes	Yes	0.05		-0.03		0.01		0.03	
	No	0.12	0.07	0.07	0.10	-0.02	-0.03	0.07	0.04

Notes & Source: See Table 1. Table displays medians of change/decade in achievement levels in math and reading for different SES constructs by subgroups and survey displayed in Tables A.20-A.27. E. A. = Early Adolescence. Bir yr = Birth year, Pared. = Parental education, Grad. = Graduated, coll. = college, h.s. = high school, Poss. = Possession index, quart=quartile and FRL = Free or reduced lunch. Birth years differ across subgroups, depending on the availability of data (see tables and appendix for details). Differences between the base category and other categories are also displayed.

Table A.20: Change in achievement levels in math and reading at younger age by parent's education levels and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Parent's Educ.	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Grad. coll.	LTT	1968-1989	25,800	0.544**	0.26	1961-1989	37,830	0.451**	0.16
				(0.020)				(0.018)	
Post h.s.			5,060	0.399**	0.19		19,000	-0.380**	-0.14
				(0.050)				(0.054)	
Grad. h.s.			11,860	0.202**	0.10		24,050	-0.123**	-0.04
				(0.036)				(0.027)	
< h.s.			3,850	0.596**	0.28		8,630	0.168**	0.06
				(0.065)				(0.055)	
Grad. coll.	NAEP	1980-1993	101,260	0.787**	0.61	1980-1993	111,090	-0.072**	-0.06
				(0.008)				(0.008)	
Post h.s.			16,060	0.735**	0.57		18,440	-0.028	-0.02
				(0.022)				(0.020)	
Grad. h.s.			22,420	0.763**	0.59		23,660	0.113**	0.09
				(0.020)				(0.018)	
< h.s.			7,750	0.811**	0.62		8,000	0.073*	0.06
				(0.032)				(0.029)	

Notes & Source: See Table 1. Table displays change in achievement levels at younger age by parent's education levels and survey. Students report their parent's education levels. For this table, the parent's education levels were recoded in four comparable categories (graduated college, post high school, graduated high school and less than high school) across all surveys (see notes in Table A.5).

Table A.21: Change in achievement levels in math and reading at early adolescence by parent's education levels and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Parent's Educ.	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Grad. coll.	LTT	1964-1998	40,410	0.433**	0.13	1957-1998	45,280	0.759**	0.19
				(0.015)				(0.016)	
Post h.s.			15,080	0.325**	0.10		29,780	-0.187**	-0.05
				(0.028)				(0.031)	
Grad. h.s.			26,100	0.247**	0.07		41,110	-0.182**	-0.04
				(0.027)				(0.026)	
< h.s.			8,850	0.502**	0.15		15,150	0.097*	0.02
				(0.045)				(0.045)	
Grad. coll.	NAEP	1976-2003	603,560	0.668**	0.25	1976-2003	661,970	0.276**	0.10
				(0.004)				(0.003)	
Post h.s.			203,910	0.470**	0.17		231,080	0.171**	0.06
				(0.007)				(0.006)	
Grad. h.s.			216,380	0.542**	0.20		235,590	0.219**	0.08
				(0.007)				(0.006)	
< h.s.			90,140	0.731**	0.27		95,870	0.356**	0.13
				(0.010)				(0.009)	
Grad. coll.	TIMSS	1985-2001	18,470	0.124**	0.08				
				(0.020)					
Post h.s.		1981-2001	11,850	0.095**	0.05				
				(0.031)					
Grad. h.s.			9,790	0.323**	0.16				
				(0.026)					
< h.s.			6,450	0.410**	0.21				
				(0.035)					

Table A.21 (Cont'd): Change in achievement levels in math and reading at early adolescence by parent's education levels and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Parent's Educ.	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Grad. coll.	PISA	1985-2000	11,660	-0.235**	-0.16	1985-2000	10,190	-0.161**	-0.11
				(0.026)				(0.024)	
Post h.s.			4,970	-0.174**	-0.12		4,620	0.012	0.01
				(0.035)				(0.036)	
Grad. h.s.			9,630	0.012	0.01		7,800	0.126**	0.08
				(0.032)				(0.032)	
< h.s.			2,160	0.309**	0.21		1,810	0.545**	0.36
				(0.062)				(0.061)	

Notes & Source: See Table 1 and Table A.20.

Table A.22: Change in achievement levels in math and reading at older age by parent's education levels and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Parent's Educ.	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Grad. coll.	LTT	1961-1995	36,230	0.112**	0.03	1954-1995	40,000	0.530**	0.13
				(0.017)				(0.016)	
Post h.s.			18,090	0.039	0.01		35,340	-0.334**	-0.08
				(0.025)				(0.023)	
Grad. h.s.			24,440	0.017	0.01		40,080	-0.242**	-0.06
				(0.030)				(0.023)	
< h.s.			10,350	0.363**	0.11		18,000	0.082*	0.02
				(0.040)				(0.036)	
Grad. coll.	NAEP	1988-1998	34,520	0.064**	0.06	1973-1998	67,470	0.082**	0.03
				(0.013)				(0.016)	
Post h.s.			15,410	0.002	0.00		33,240	-0.049**	-0.02
				(0.017)				(0.016)	
Grad. h.s.			12,770	0.021	0.02		27,180	-0.039*	-0.02
				(0.020)				(0.019)	
< h.s.			5,620	0.084**	0.08		11,520	0.098**	0.04
				(0.029)				(0.026)	

Notes & Source: See Table 1 and Table A.20.

Table A.23: Change in achievement levels in math and reading at younger age by possession index quartiles and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Quartile	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Top	LTT	1968-2002	88,570	0.737**	0.22	1961-2002	116,440	0.127**	0.03
				(0.021)				(0.021)	
Second				0.665**	0.20			0.130**	0.03
				(0.022)				(0.021)	
Third				0.716**	0.21			0.195**	0.05
				(0.022)				(0.022)	
Bottom				0.999**	0.29			0.410**	0.10
				(0.022)				(0.022)	
Top	NAEP	1980-2007	1,408,140	0.943**	0.35	1980-2007	1,557,550	0.131**	0.05
				(0.005)				(0.004)	
Second				1.168**	0.43			0.281**	0.10
				(0.005)				(0.004)	
Third				1.073**	0.40			0.236**	0.09
				(0.005)				(0.004)	
Bottom				1.084**	0.40			0.291**	0.11
				(0.005)				(0.004)	
Top	TIMSS	1985-2005	51,020	0.624**	0.31				
				(0.022)					
Second				0.555**	0.28				
				(0.021)					
Third				0.526**	0.26				
				(0.020)					
Bottom				0.716**	0.36				
				(0.020)					

Table A.23 (Cont'd): Change in achievement levels in math and reading at younger age by possession index quartiles and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Quartile	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Top	PIRLS					1991-2006	25,840	0.154**	0.10
								(0.029)	
Second								0.130**	0.09
								(0.028)	
Third								0.035	0.02
								(0.029)	
Bottom								0.170**	0.11
								(0.030)	

Notes & Source: See Table 1. Table displays change in achievement levels at younger age by home item quartiles and survey. Students report items in their homes. For this table, we included items that were consistently available and similarly measured for all administrations within a subject, age/grade, and survey. (see notes in Table A.6).

Table A.24: Change in achievement levels in math and reading at early adolescence by possession index quartiles and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Quartile	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Top	LTT	1964-1998	99,920	0.578**	0.17	1957-1998	120,820	0.246**	0.06
				(0.021)				(0.022)	
Second				0.391**	0.12			0.008	0.00
				(0.021)				(0.022)	
Third				0.472**	0.14			0.084**	0.02
				(0.022)				(0.024)	
Bottom				0.499**	0.15			0.141**	0.03
				(0.022)				(0.025)	
Top	NAEP	1976-2003	1,259,180	0.770**	0.29	1976-2003	1,380,970	0.368**	0.14
				(0.005)				(0.005)	
Second				0.759**	0.28			0.364**	0.13
				(0.006)				(0.005)	
Third				0.619**	0.23			0.200**	0.07
				(0.006)				(0.005)	
Bottom				0.719**	0.27			0.366**	0.14
				(0.006)				(0.005)	
Top	TIMSS	1981-2001	56,260	0.351**	0.18				
				(0.022)					
Second				0.343**	0.17				
				(0.021)					
Third				0.303**	0.15				
				(0.021)					
Bottom				0.488**	0.24				
				(0.020)					

Table A.24 (Cont'd): Change in achievement levels in math and reading at early adolescence by possession index quartiles and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Quartile	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Top	PISA	1985-2000	28,800	-0.315**	-0.21	1985-2000	24,860	-0.214**	-0.14
				(0.030)				(0.029)	
Second				-0.224**	-0.15			-0.136**	-0.09
				(0.030)				(0.029)	
Third				-0.164**	-0.11			-0.006	0.00
				(0.030)				(0.029)	
Bottom				0.114**	0.08			0.234**	0.16
				(0.030)				(0.030)	

Notes & Source: See Table 1 and Table A.23.

Table A.25: Change in achievement levels in math and reading at older age by possession index quartiles and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Math				Reading			
Quartile	Survey	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Top	LTT	1961-1995	89,470	0.313**	0.09	1954-1995	113,910	0.164**	0.04
				(0.024)				(0.021)	
Second				0.125**	0.04			-0.051*	-0.01
				(0.024)				(0.021)	
Third				0.138**	0.04			-0.110**	-0.03
				(0.025)				(0.022)	
Bottom				0.224**	0.07			0.023	0.01
				(0.025)				(0.024)	
Top	NAEP	1988-1998	70,740	0.189**	0.19	1973-1998	143,960	0.264**	0.11
				(0.018)				(0.015)	
Second				0.155**	0.16			0.174**	0.07
				(0.017)				(0.015)	
Third				-0.261**	-0.26			-0.080**	-0.03
				(0.018)				(0.016)	
Bottom				0.078**	0.08			-0.072**	-0.03
				(0.017)				(0.017)	

Notes & Source: See Table 1 and Table A.23.

Table A.26: Change in achievement levels in math and reading by eligibility for free or reduced lunch and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Math				Reading			
FRL elig.	Survey	Grade/Age	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
No	LTT	Age 9	1994-2002	11,080	0.212**	0.27	1994-2002	11,180	0.187**	0.23
					(0.021)				(0.019)	
Yes				10,970	0.182**	0.23		10,830	0.187**	0.23
					(0.023)				(0.020)	
No	NAEP	Grade 4	1986-2007	669,150	0.747**	0.36	1988-2007	744,280	0.226**	0.12
					(0.004)				(0.003)	
Yes				705,110	0.800**	0.38		759,620	0.271**	0.14
					(0.004)				(0.003)	
No	LTT	Age 13	1990-1998	12,570	0.222**	0.28	1990-1998	12870	0.277**	0.35
					(0.018)				(0.022)	
Yes				9,490	0.205**	0.26		9,770	0.185**	0.23
					(0.021)				(0.024)	
No	NAEP	Grade 8	1982-2003	655,400	0.535**	0.25	1984-2003	716,580	0.236**	0.12
					(0.004)				(0.003)	
Yes				566,760	0.557**	0.27		599,240	0.204**	0.11
					(0.005)				(0.004)	
No	LTT	Age 17	1987-1995	14,420	0.138**	0.17	1987-1995	14,950	0.194**	0.24
					(0.017)				(0.019)	
Yes				7,620	0.104**	0.13		7,950	0.230**	0.29
					(0.025)				(0.026)	
No	NAEP	Grade 12	1988-1998	45,140	0.162**	0.16	1981-1998	65,960	0.088**	0.05
					(0.011)				(0.012)	
Yes				21,230	0.164**	0.16		29,420	0.052**	0.03
					(0.016)				(0.019)	

Notes & Source: See Table 1. Table displays change in achievement levels eligibility for free or reduced lunch and survey.

Table A.27: Change in achievement levels in math and reading by family structure and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Subject		Math				Reading			
#Parents	Survey	Grade/Age	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Two	LTT	Age 9	1976-1989	28,950	0.264**	0.20				
					(0.016)					
Other				8,470	0.300**	0.23				
					(0.030)					
Two	NAEP	Grade 4	1980-2007	365,080	0.995**	0.37	1980-2007	369,400	0.144**	0.05
					(0.004)				(0.003)	
Other				118,300	1.104**	0.41		118,160	0.329**	0.12
					(0.007)				(0.006)	
Two	LTT	Age 13	1972-1985	25,400	0.189**	0.15				
					(0.014)					
Other				7,360	0.056*	0.04				
					(0.027)					
Two	NAEP	Grade 8	1976-2003	334,850	0.640**	0.24	1976-2003	341,210	0.259**	0.10
					(0.004)				(0.004)	
Other				124,320	0.587**	0.22		124,050	0.281**	0.10
					(0.007)				(0.006)	
Two	TIMSS	Grade 8	1981-1985	14,550	0.290**	0.73				
					(0.017)					
Other				4,430	0.329**	0.82				
					(0.031)					
Two	PISA	Age 15	1985-1997	11,130	-0.235**	-0.20	1985-1997	11,940	-0.187**	-0.16
					(0.022)				(0.020)	
Other				5,790	-0.006	-0.01		6,580	0.044	0.04
					(0.033)				(0.032)	

Table A.27 (Cont'd): Change in achievement levels in math and reading by family structure and survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Subject		Math				Reading			
#Parents	Survey	Grade/Age	Birth Cohorts	N	Change	Change/dc	Birth Cohorts	N	Change	Change/dc
Two	LTT	Age 17	1969-1982	15,030	0.153**	0.12				
					(0.020)					
Other				4,510	0.225**	0.17				
					(0.036)					
Two	NAEP	Grade 12					1973-1998	37,950	0.029*	0.01
									(0.014)	
Other								14,610	-0.038	-0.02
									(0.022)	

Notes & Source: See Table 1. Table displays change in achievement levels by family structure and survey (reported by students: two parent family and others; surveys do not ask if the parents are biological parents) (see details in Table A.7).

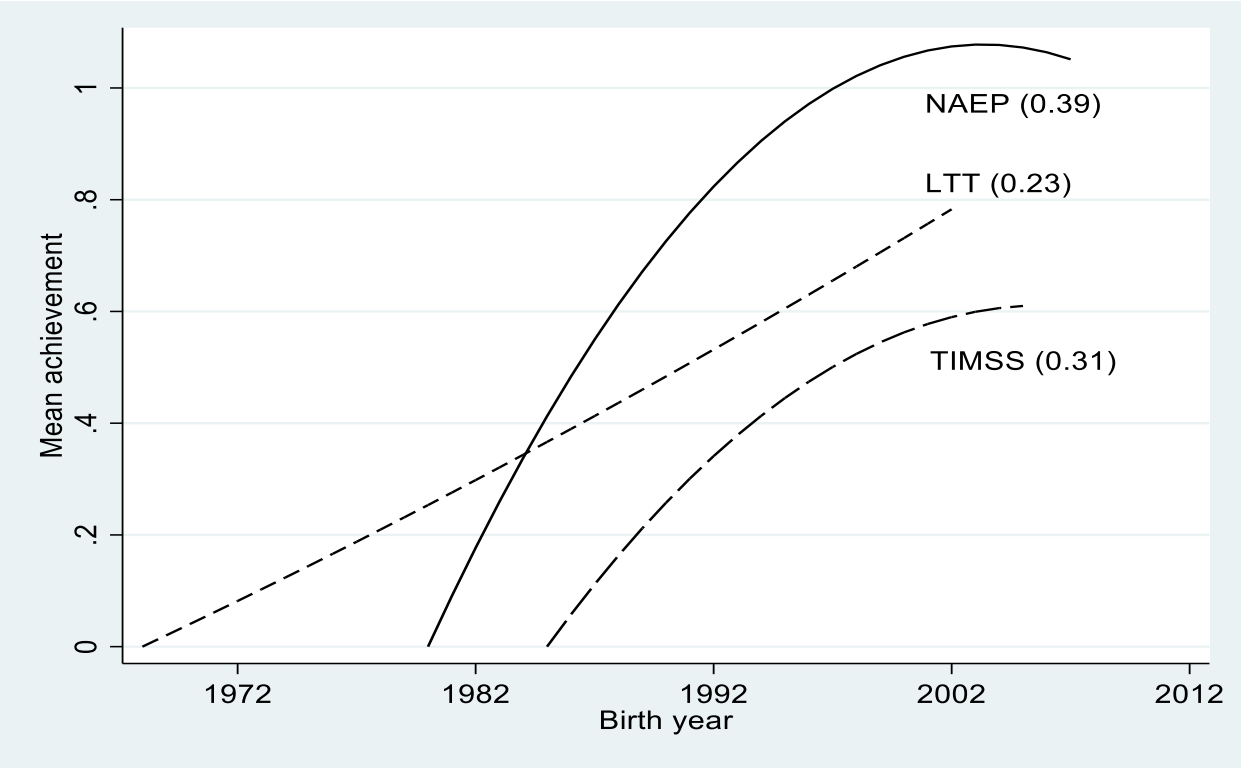

Figure A.1. Change in achievement levels in math at younger age by survey: Birth Cohort 1968-2007.
Note: Figure shows change in achievement levels at younger age by survey. Normalized achievement is measured in standard deviations (s.d.). The s.d. is the difference between the year test was administered and the starting year for a specific test series. The lines represent a quadratic fit. Each line is forced to begin at zero. Magnitude of the rate of change in achievement levels per decade is displayed in parenthesis for each line.
Source: See Table 1.

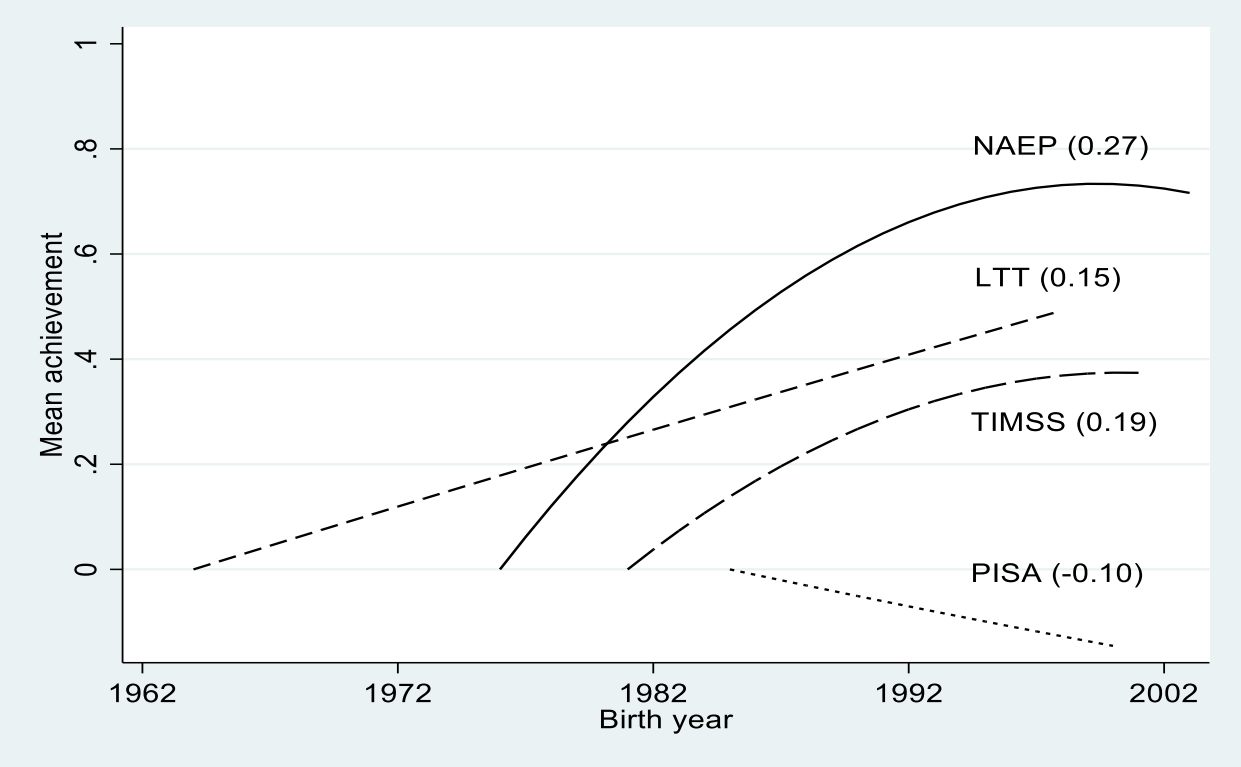

Figure A.2. Change in achievement levels in math at early adolescence by survey: Birth Cohort 1964-2003.
Notes & Source: See Figure A.1 and Table 1.

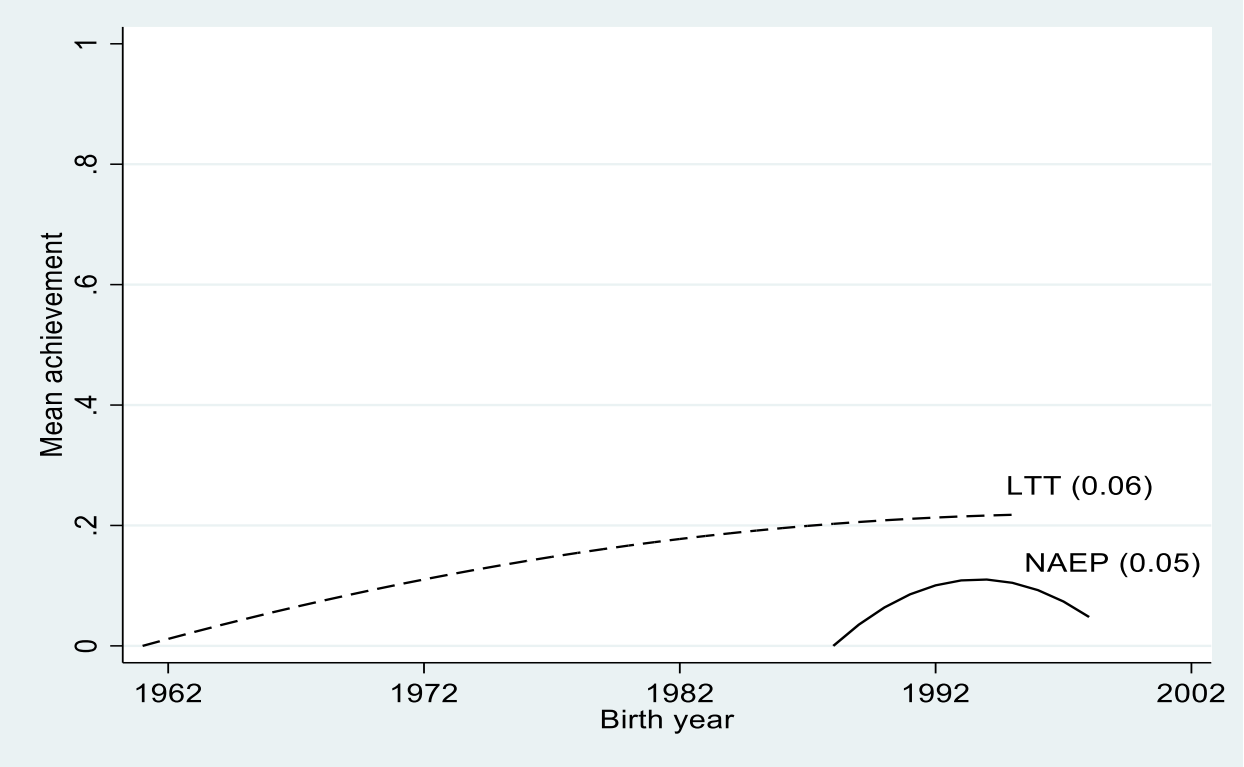

Figure A.3. Change in achievement levels in math at older age by survey: Birth Cohort 1961-1998.
Notes & Source: See Figure A.1 and Table 1.

Figure A.4. Change in achievement levels in reading at young age by survey: Birth Cohort 1961-2007.
Notes & Source: See Figure A.1 and Table 1.

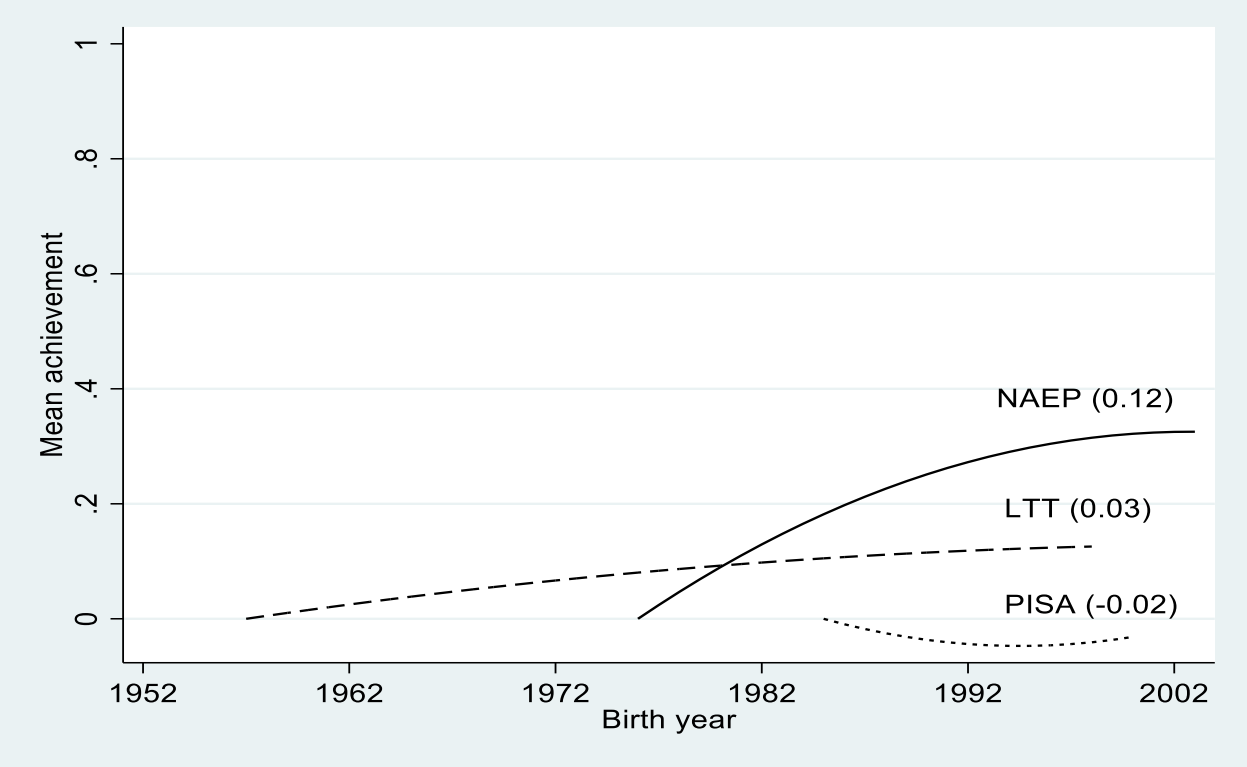

Figure A.5. Change in achievement levels in reading at early adolescence by survey: Birth Cohort 1957-2003.
Notes & Source: See Figure A.1 and Table 1.

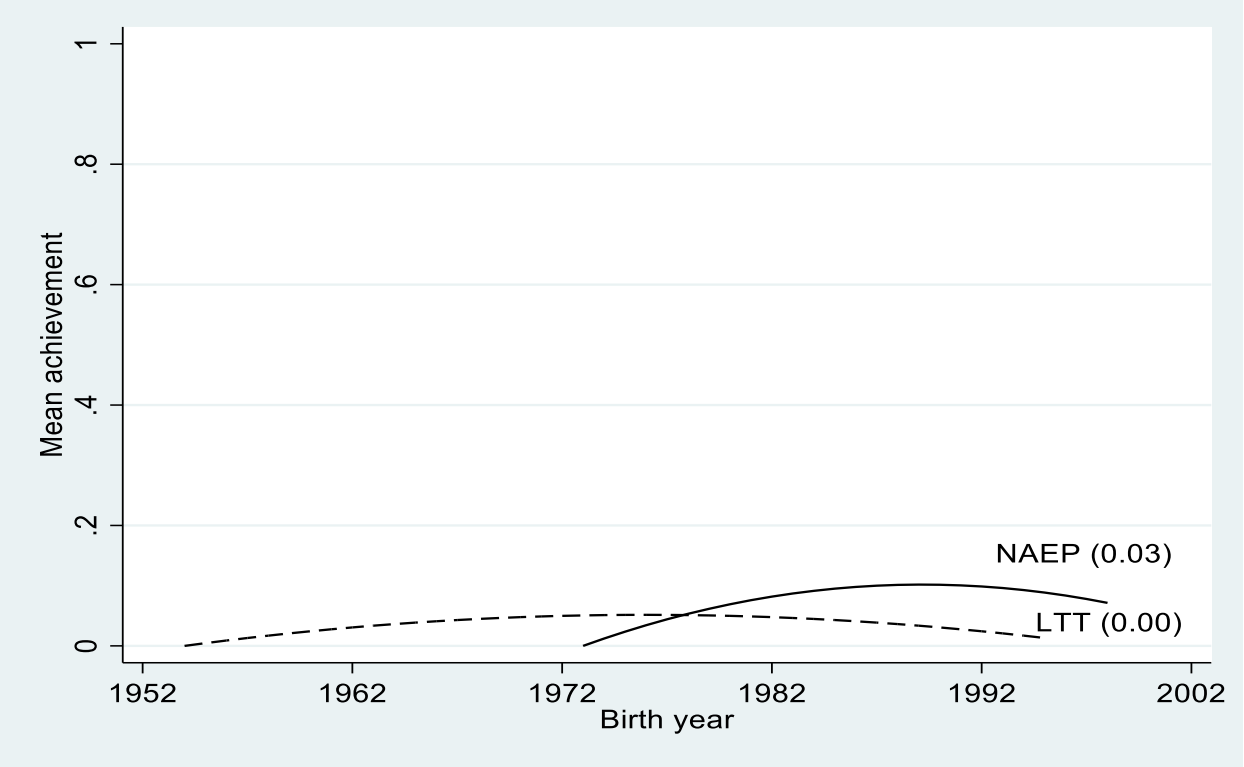

Figure A.6. Change in achievement levels in reading at older age by survey: Birth Cohort 1954-1998.
Notes & Source: See Figure A.1 and Table 1.

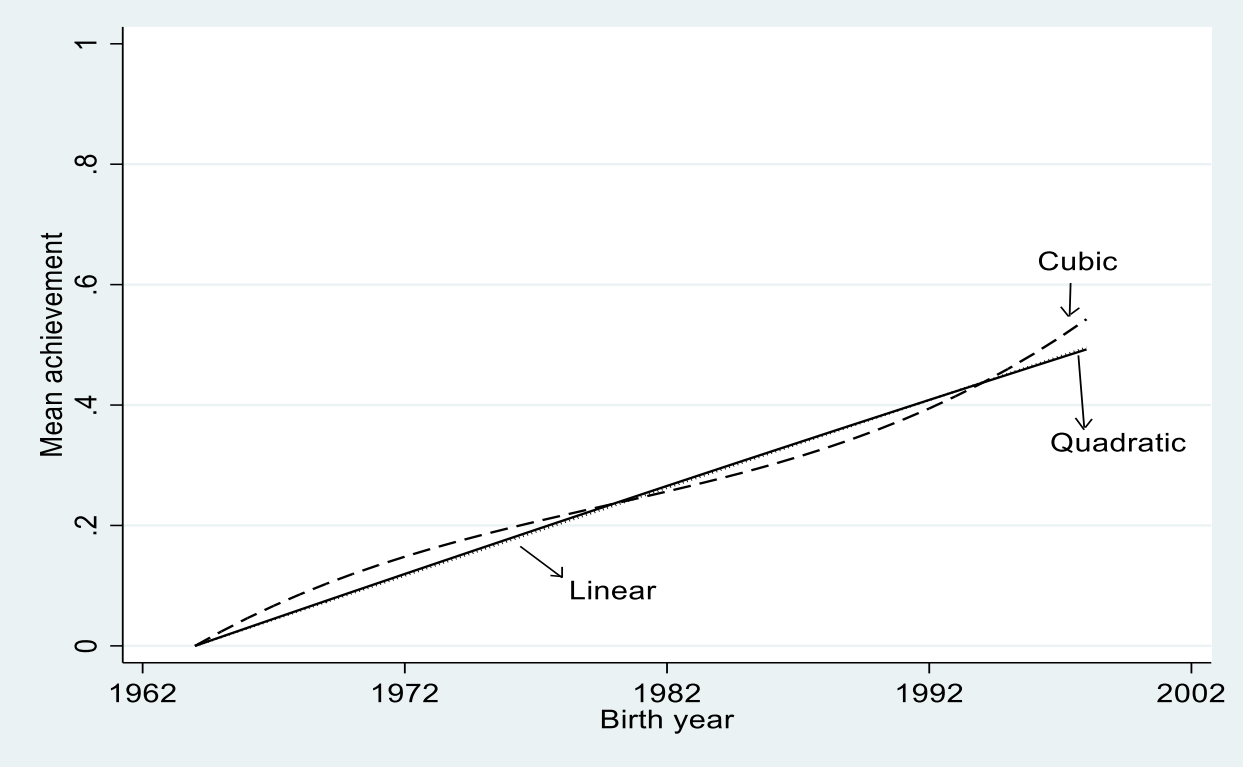

Figure A.7. Change in achievement levels in math at age 13 in LTT for linear, quadratic, and cubic fit. These changes are 0.50, 0.49 and 0.54 standard deviations.
Notes & Source: See Figure A.1 and Table 1.

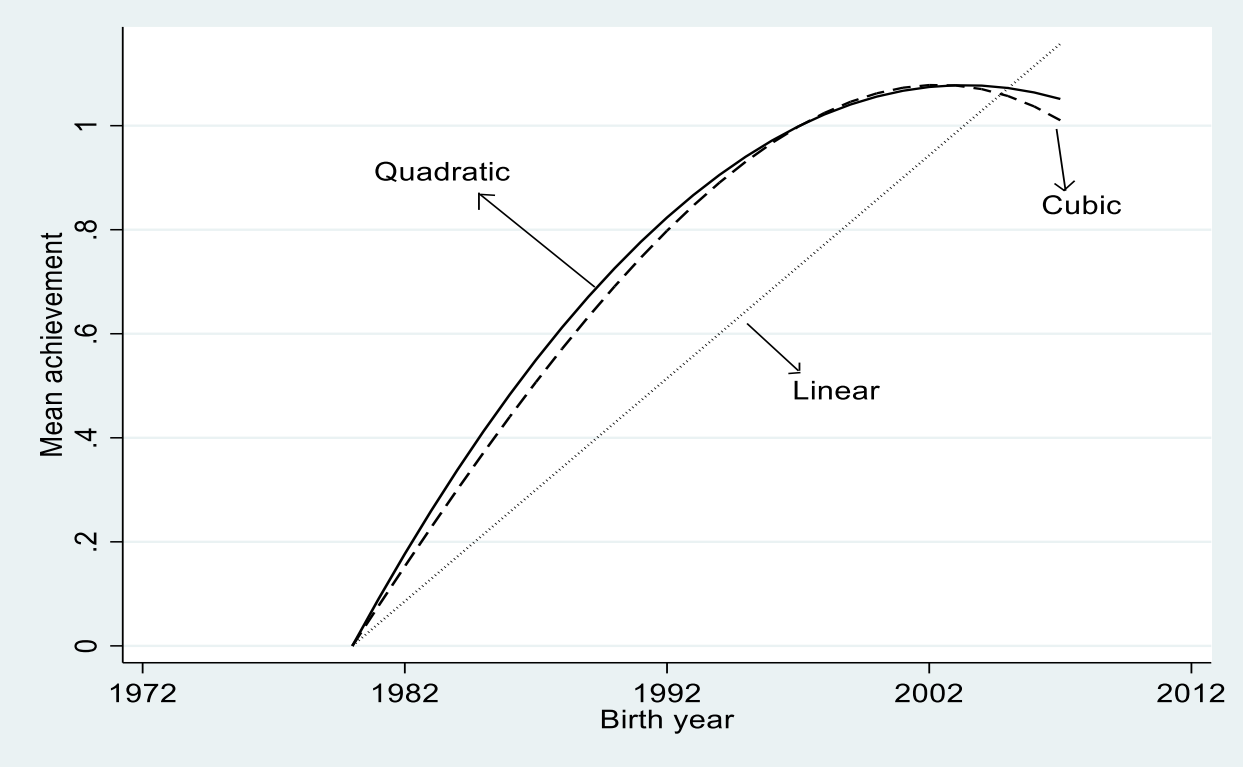

Figure A.8. Change in achievement levels in math at grade 4 in NAEP for linear, quadratic, and cubic fit. These changes are 1.16, 1.05 and 1.01 standard deviations.
Notes & Source: See Figure A.1 and Table 1.

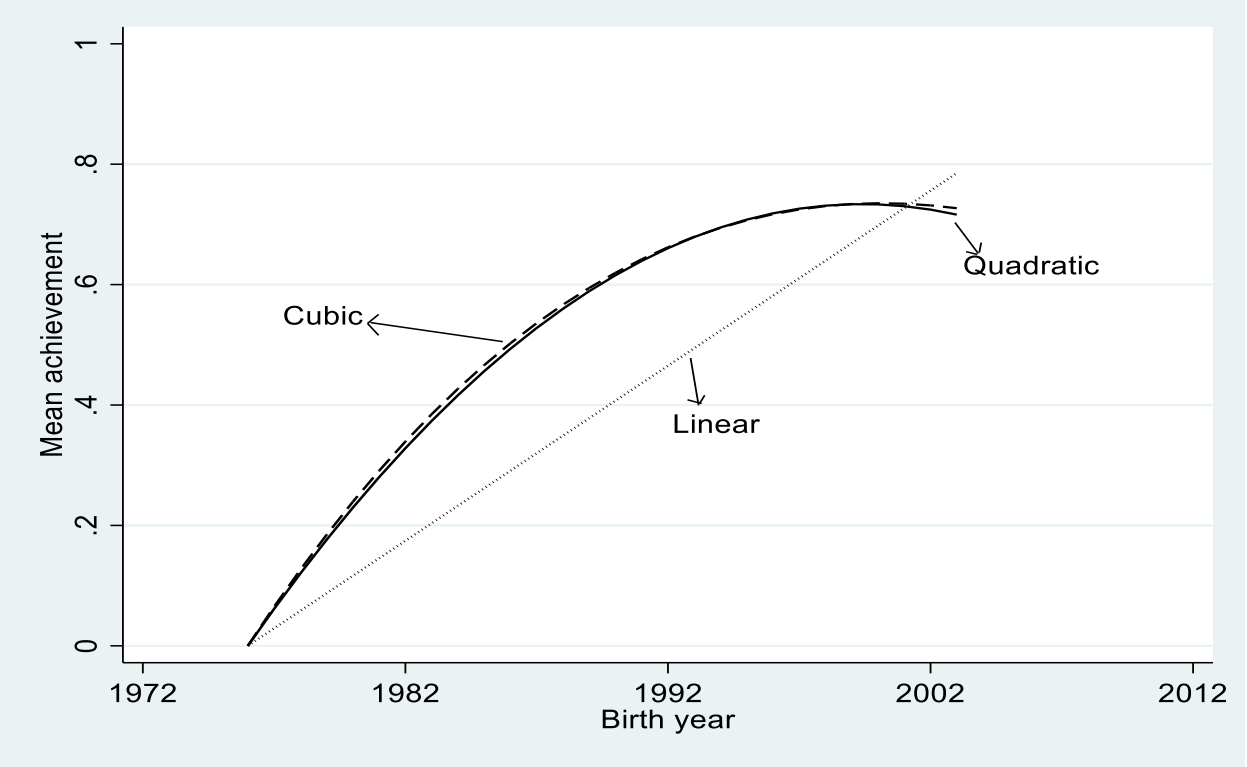

Figure A.9. Change in achievement levels in math at grade 8 in NAEP for linear, quadratic, and cubic fit. These changes are 0.79, 0.72 and 0.73 standard deviations.
Notes & Source: See Figure A.1 and Table 1.