Notes on Trends in the U.S. Income-Achievement Gap

Eric A. Hanushek, Jacob Light, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann*

June 2022

* Hanushek: Hoover Institution/Stanford University, NBER, and CESifo; hanushek@stanford.edu. Light: Stanford University; jdlight@stanford.edu. Peterson: Harvard University and Hoover Institution/Stanford University; ppeterso@gov.harvard.edu. Talpey: Hoover Institution/Stanford University; lauratalpey@gmail.com. Woessmann: University of Munich, ifo Institute, and CESifo; woessmann@ifo.de.

Reardon (2011b) reports rapidly growing SES gaps when SES is measured by current parental income indicators. This differs dramatically from our analysis of trends in the SES-achievement gap (Hanushek et al. (2022)). We reconcile our findings with his by showing that the difference in results is due to systematic measurement errors in his analysis.[1] When a trend line is estimated from the studies in his analysis less prone to measurement error, no upward trend in the SES-achievement gap is detected.

Reardon estimates trends in the income-achievement gap for cohorts born between 1942 and 2001 from observations of average math and reading test performance obtained from "twelve nationally representative" samples of students of various ages. Table 1 provides a list of the included databases and their acronyms. The underlying databases were selected by Reardon because they were accompanied by surveys that included student- or parent-provided data on household income (p. 93) along with student achievement test information.

Within each data set, Reardon computes a value of the achievement gap between students at the 90th percentile of the income distribution and those at the 10th percentile from a cubic regression of achievement on parental income.[2] He treats every age cohort within each test administration as an independent observation, giving him 20 estimates of math performance gaps and 26 of reading performance gaps. These are the data points for the subsequent trend analysis.

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

Reardon (2011b) himself expresses concern "that the trend in the estimated gaps for the earliest cohorts, those born before 1970, is not as accurately estimated as the later trend. … Family income was reported by students rather than by a parent … [and they] are school-based samples of students in high school [that] exclude dropouts, who are disproportionately low-

[1] The studies differ in several ways including the measurement of SES, the parts of the SES distribution that are compared, and choice of underlying data used for the trends. As described below, we believe that the last issue – the choice of underlying data sources – is the dominant factor in the differing results in trend lines.

[2] Issues related to extrapolating the observed SES percentiles are discussed in Hanushek et al. (2022) and present an additional component of measurement error to that emphasized in this note.

income and low-achieving" (pp. 95-96). He is more confident that the trend has shifted upward for cohorts born between 1970 and 2001. Between 1974 and 2001, "the income achievement gap has grown by roughly 40 to 50 percent …, a very sizeable increase" (p. 97). Elsewhere, he describes this increase as "roughly 30 to 40 percent" (p. 93). Unfortunately, his analysis does not reflect the overall impact of the data problems, many of which he himself identifies in an appendix (Reardon (2011a)).

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

Data Quality Issues

Survey researchers have not found it easy to collect accurate data on family income, the independent variable in the regressions that Reardon uses to create the gap estimates for each survey. The challenge is especially large when household income is estimated from student-provided information (Kayser and Summers (1973); Fetters, Stowe, and Owings (1984); Kaufman and Rasinski (1991)). Surveys that do not get information from parents typically rely on questions to students about their parents' education and household possessions in order to develop measures of SES.

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

Six of his twelve surveys are plagued with serious problems and clearly do not meet current scientific quality standards. Importantly, these surveys introduce systematic measurement error into the subsequent trend estimation based upon them. From the information provided in Reardon (2011a, (2011b) and the surveys self-descriptions available online, these surveys do not provide reliable data points but nonetheless have a decided influence on the shape of the estimated trend line.

Project Talent

Talent, the earliest survey, provides estimated gaps for four cohorts born in 1942-1945. It uses student-provided estimates of family income, employing five income categories.[3] While there are questions about the overall sampling in the Talent survey, the largest concerns relate to nonresponse rates to the question concerning family income. No less than 54 percent of the freshmen, 50 percent of the sophomores, 45 percent of the juniors, and 39 percent of the seniors chose not to "guess"—the word used in the survey—the answer to the family income question.

For the early Talent cohorts born in the early forties, Reardon reports income-achievement gaps in reading and math of approximately 0.75 s.d.[4] Only Prospect, which has its own measurement problem (see below), reports gaps of a similarly low magnitude.

National Longitudinal Study of 1972 (NLS-72)

The National Longitudinal Study of 1972 was an early survey conducted by the National Center for Education Statistics. It was designed to follow a sample of high-school seniors of the Class of 1972 into the labor market and college. Like Talent, its parental-income measure came from sampling the students, who place estimated income into one of ten categories.[5] Twenty-one percent of the respondents chose not to respond to this question.

[3] The impact of having just five categories depends on where these categories cut the income distribution. Reardon (2011a) considers the impact of differing numbers of categories that are evenly spaced across the income distribution. Actual survey data on categorical income, however, is almost certainly going to have unevenly spaced data categories with limited observations in the tails of the distribution – exactly where he focuses his analysis.

[4] We estimate the size of the gaps from Reardon, 2011, Figures 5.3 and 5.4, pp. 98-99, which are too small to allow for precise estimation. No table showing the precise size of the estimated gap is available either in the paper or its Appendix.

[5] The top household income category in the NLS-72 survey was income greater than $18,000. From the relevant Current Population Survey (CPS), 16 percent of U.S. households had incomes that would be included in this top category.

High School and Beyond (HS&B)

The High School and Beyond study collected income data for birth cohorts in 1962 and 1964. An important element of these surveys was a 15 percent sample that collected income data from both a parent and the student. These data were particularly important in the analysis because they provided estimates of the impact of student-provided income data. Unfortunately, the parent-provided income data lacked face validity, being considerably above the comparable values from the Current Population Survey (CPS).[6] Reardon uses these parent-provided data to estimate reliabilities for the student-provided data. He concludes that the student data are more reliable than the parent data. Partly because of this and partly because of sample sizes, he uses the student-provided data from HS&B to estimate the income-achievement relationship and the subsequent SES-gap points for the trend analysis.

To get some sense of the impact that measurement error can have on estimates of the income-achievement gap, one can compare HS&B estimates to those observed in NLSY79, which was administered to birth cohorts of students that overlapped with those observed in HS&B.[7] NLSY79 is much less likely than HS&B to err in its estimation of household income. The highly-regarded NLSY79 protocol, administered to parents, obtains household income as a continuous variable with just a 2 percent non-response rate. In contrast, the student-provided HS&B data are categorical and have non-response rates of 13 and 18 percent of seniors and sophomores, respectively. Because of the differential measurement error, it is not surprising that the average of the math and reading income-achievement gaps identified in the NLSY79 survey is 1.3 s.d., which is fully 0.35 s.d. higher than the gap reported in the HS&B survey.

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

Problems with the seldom-used Prospects survey are noted in Reardon (2011b), p. 112, note 70: "It is difficult to find documentation on the content and psychometric properties of the Prospects tests. These tests may be much less reliable than other tests; as a result, I am inclined

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

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

to discount their importance in describing the trends." Contrary to that statement, Reardon uses three observations from Prospects to estimate trend lines in both math and reading, while six other studies yield only one observation.[8]

The estimates of the income-achievement gap from Prospects vary widely across the three cohorts surveyed at the same time—by 0.3 s.d. in math and 0.35 s.d. in reading. All three of the Prospects estimates are much lower than the 1.3 s.d. gap obtained from NLSY97 administered at about the same time. In other words, estimates for the income-achievement gap for cohorts born within a couple of years of 1980 vary by no less than 0.6 s.d. Measurement error is the most likely explanation for this spread in estimates made at close to the same time.

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

Add Health contains no math achievement data, but the study contributes over a third of the post-1970 data points used by Reardon to estimate the trend in the reading gap. The observations cover students at the ages of 13 through 18 born between 1978 and 1983. The survey has been used for a wide variety of studies, but grades and GPA are the primary school outcome measures. As Rees and Sabia (2010) note, "the Adolescent Health study did not administer formal achievement tests such as are available in the National Education Longitudinal Survey of 1988." Instead, it contains the Peabody Picture Vocabulary Test (PPVT), a test of receptive vocabulary. It is described on the NLSY79 website (which used this test along with reading and math achievement assessments) as follows: "The PPVT-R [an updated version of the original] consists of 175 stimulus words and 175 corresponding image plates. Each image plate contains 4 black-and-white drawings, one of which best represents the meaning of the corresponding stimulus word. … Starting in 1998, the administration of the PPVT-R was largely limited to 4- and 5-year-old children".[9]

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

[8] Only one birth cohort is observed with data from NLS-72, NELS, ELS, SECCYD, ECLS-K, and ECLS-B.

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

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

Study of Early Child Care and Youth Development (SECCYD)

Although Reardon says his surveys are nationally representative, SECCYD is not. According to U.S. Department of Health and Human Services (2018), the study recruited ten hospitals willing to participate as recruiting grounds for mothers willing and available to participate in a multi-year study. The sample description explicitly says its "data are not representative in the statistical sense, and therefore inference to the nation as a whole is not possible. Comparisons to other databases, national or otherwise, should be made with extreme caution."[10]

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

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

It is natural to want to include as many different survey data sets as possible. But, as delineated above, six of Reardon's choices do not meet current scientific quality standards. The three early surveys rely upon student reports of household income (Talent, NLS-72, and HS&B); three later surveys lack reliable test information and/or pertain to non-representative samples (Prospects, Add Health, and SECCYD). These error-prone data points yield gap estimates that are biased downward. Since they tend to relate to early birth cohorts, they distort the trends that are estimated when combined with higher-quality data available later.

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

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

achievement gap analysis to include the ECLS-K2010.[11] (The ECLS-K2010 observation became available subsequent to the publication of Reardon (2011b) and extends the birth cohorts available to 2005).

Figure 1 replicates the Reardon analysis of estimated 90-10 income-achievement gaps in math and reading for the birth cohorts 1961-2005 using only the high-quality surveys.[12] All of these data come from well-regarded, nationally representative surveys that obtained income information from parents rather than students. One striking fact emerges from this figure: Simple linear regressions that estimate trends in math and reading from the ten observations taken from these surveys show *perfectly flat trends* with no significant change in the income-achievement gap over this time period (Figure 1).

Trends from Two Sets of Psychometrically Linked Surveys

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

It is possible to use two subsets of Reardon's data to investigate the potential impact of testing artifacts on the trend results. The first set of surveys relying on the same testing, NLSY79 and NLSY97, was administered to cohorts born as early as 1961 and as recently as 1981, a twenty-year interval.[13] The second set, ECLS-K, ECLS-B, and ECLS-K2010, provides information on income and achievement for cohorts born as early as 1993 and as late as 2005, a twelve-year interval. Together, these two sets of studies overlap nearly half (28 of 59 years) of the entire time span covered by Reardon (2011b) and over half (15 of 26 years) of the period between 1974 and 2001 when Reardon is more confident of the accuracy of estimated gap growth, identified as at least "30 to 40 percent".

[11] While the three ECLS data sets are included among the high-quality surveys, their outcome data are tests of kindergarten readiness in math and reading. Reardon and Portilla (2016), who specifically analyze trends in these data, argue that these SES-achievement gaps are comparable to those for later ages in the other surveys because gaps do not change much over the school years.

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

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

If the income gap had truly increased, it should be evident in these two intertemporally linked surveys. But, at roughly 1.3 s.d., the average income-achievement gap in math and reading estimated by NLSY97 is no different from the one observed twenty years earlier. Over the twelve years between the administration of ECLS-K and ECLS-K2010, the 90-10 math achievement gap, as reported by Reardon and Portilla (2016), declines by 0.13 s.d. and the gap for reading drops by 0.21 s.d.[14]

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

Conclusion

Both estimates from all high-quality surveys (Figure 1) and those from psychometrically matched surveys indicate no substantial change in the income-achievement gap in math and reading over time. The size of the gap remains constant at roughly the same level as the SES-achievement gap reported in this paper, indicating that analyses based on different surveys and different underlying SES measures actually come to the same broad conclusion. In other words, the apparently contradictory results between our analysis of SES-achievement trends and those of Reardon (2011b) are completely resolved by recognition of the systematic measurement error embedded in specific surveys used in Reardon.[15]

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

[15] Importantly, the apparently different trend results of Reardon (2011b) and Hanushek et al. (2022) do not appear to arise from other analytical differences. They do not reflect the specific measure of SES (current family income versus a composite of parental education and items in the home). Nor do they reflect a focus on specific percentiles of the SES distribution (90-10) versus aggregate gaps in the tails of the distribution (75-25).

References

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

Hanushek, Eric A., Jacob Light, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. 2022. "Long-run Trends in the U.S. SES-Achievement Gap." Stanford, CA: Hoover Insitution (June).

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

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

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

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

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

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

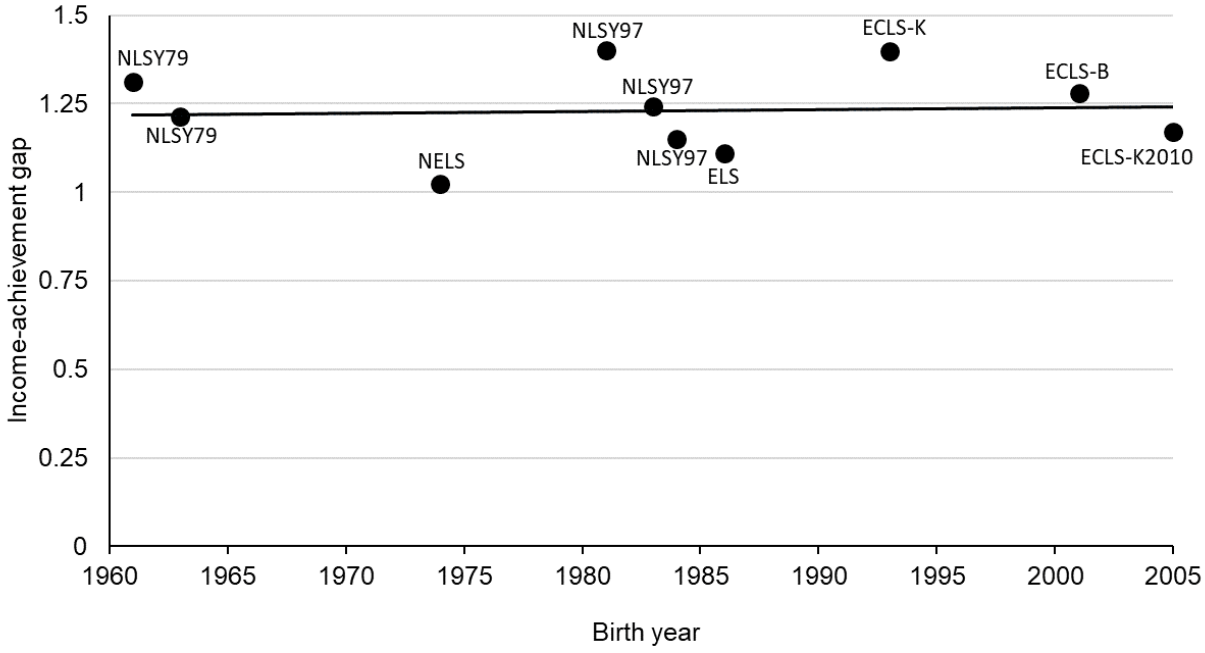
Reardon, Sean F., and Ximena A. Portilla. 2016. "Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry." *AERA Open* 2, no. 3: 1-18.

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

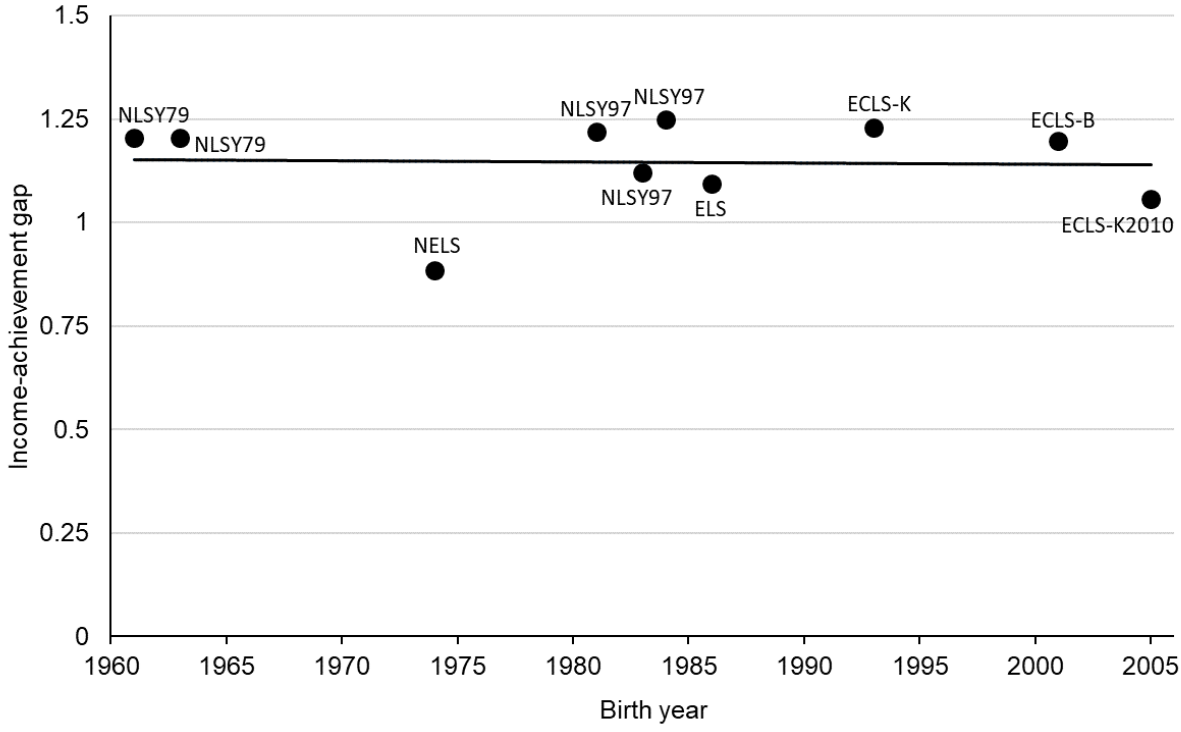
U.S. Department of Health and Human Services. 2018. NICHD Study of Early Child Care and Youth Development: Phase I, 1991-1994: Inter-university Consortium for Political and Social Research.

Figure 1. Estimated 90-10 Income-Achievement Gaps from High-quality Surveys, Birth Cohorts 1961-2005

Panel A: Math



Panel B: Reading



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

Table 1. Acronyms for Sources of Data Analyzed by Reardon (2011b)

Acronym	Survey
Talent	Project Talent
NLS-72	National Longitudinal Study of 1972
HS&B	High School and Beyond
NLSY79	National Longitudinal Study of Youth 1979
NELS	National Education Longitudinal Study
Prospects	The Congressionally Mandated Study of Educational Growth and Opportunity
Add Health	National Longitudinal Study of Adolescent to Adult Health
NLSY97	National Longitudinal Study of Youth 1997
ELS	Education Longitudinal Study
SECCYD	Study of Early Child Care and Youth Development
ECLS-K	Early Childhood Longitudinal Study-Kindergarten, 1997
ECLS-B	Early Childhood Longitudinal Study, Birth Cohort