Appendices for:

The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations

sean f. reardon

Stanford University
520 Galvez Mall, # 526
Stanford, CA 94305
650.736.8517
sean.reardon@stanford.edu

5.A1. Studies Used

The studies used are listed in Table 5.A1.

Table 5.A1 here

5.A2. Computing Achievement Gaps

To standardize the test scores in each study used here, I first fit a regression model

$$Y_i = \beta_0 + \beta_1(AGE_i) + e_i, \quad e_i \sim N(0, \sigma^2),$$

using the appropriate sample weights. This yields an estimate of the age-adjusted variance in test scores, $\hat{\sigma}^2$, and an estimated residual, $\hat{e}_i$, for each student. Dividing the residual by the root mean squared error yields the age-adjusted standardized test score for each student

$$\hat{Y}_i^* = \frac{\hat{e}_i}{\hat{\sigma}}.$$

By construction, the $\hat{Y}_i^*$'s have a mean of 0 and a standard deviation of 1 (when weighted by the appropriate sample weight).

Although standardizing test scores solves the primary problems of the comparability of gaps measured with different tests and in non-interval-scale metrics, there are several potential problems. First, suppose we have some 'true' measure of cognitive ability, measured in a meaningful interval scale. If the variance of academic achievement, as measured in this metric, changes over time (either across cohorts, or within a cohort as it progresses through school), then standardizing the metric at each wave of testing confounds changes in the 'true' gap with changes in the variance of test scores. In this chapter, this will be a problem if the true variance of academic achievement varies over time. If the true variance of academic achievement grows over time, then the estimated trend in the achievement gaps will be underestimated, and vice versa. An examination of the standard deviations of LTT-NAEP scores from the 1970s through the 2000s, however, shows no trend in their magnitude, suggesting that the true variance of academic achievement has not changed appreciably over the last 40 years (Jencks, Owens, Shollenberger, & Zhu, 2010).

Second, measurement error in test scores will tend to inflate the variance of the test score distributions (thereby inflating $\hat{\sigma}^2$), meaning that the achievement gaps measured in standard deviation units will be biased toward zero. If the gaps at different grades, ages, or cohorts are measured with tests that have different amounts of measurement error, then the amount of bias will not be the same in each measure of the gap, leading to potentially erroneous inferences regarding patterns or trends in the magnitudes of the gaps over time.[1] Table 5.A2 provides information regarding the reliability of the tests used in each of the studies. In order to correct gap estimates for measurement error, I multiply the estimates by $\frac{1}{\sqrt{r}}$, where r is the reliability of the test. This yields estimates of the true gaps, and eliminates any bias in the trend that may arise from differential reliability of the tests.

Table 5.A2 here

Note that for some studies, estimates of the reliability of the tests was not available. In particular, I could find no information on the reliability of the tests in Project Talent and EEO; for both I assumed a reliability of 0.75, which is slightly lower than the lowest-reported reliability for any of the other studies. This is a conservative choice, as lower reliabilities will inflate the estimated gap for these studies, and will thereby attenuate the estimated trend in income achievement gaps.

[1] This bias is not likely to be large, given that most standardized tests relatively high reliabilities, typically between 0.7 and 1.0. If two tests measure the same thing with different reliabilities, the ratio of the gaps estimated from the two different tests will be $\sqrt{r_1/r_2}$, where r_1 and r_2 are the reliabilities of the two tests. Thus, two tests with reliabilities of 0.7 and 1.0 would yield gap estimates that differed from one another by less than 20%.

Measuring socioeconomic achievement gaps

In general, we would like to estimate the association between student test scores and a measure of parental socioeconomic status (e.g., family income, parental education, or a composite SES score). We would like this measure to be comparable across studies conducted in different years and that collect information on income and parental education differently. Because dollars are not comparable across years, and because income inequality changes over time, we require a method that characterizes the relationship between income and children's achievement in a way that is comparable over time. I use the 90/10 income achievement gap—the difference in the average standardized test scores of children at the 90th percentile of children's family income and of children at the 10th percentile of family income.

Because income is generally measured categorically, in 5-15 ordered income categories, we cannot identify children's exact income or their exact percentile in the income distribution. Below I describe a method for estimating the average test score of children at any given percentile of the income distribution based on categorical income data. I begin with a general formulation of the problem and an approach to addressing it, and then describe the specific approach I take in this chapter.

Suppose there is a continuous latent family trait θ that is distributed according to the density function $\phi(\theta)$ (with cumulative density function $\Phi(\theta)$) in the population. We observe a crudely measured version of θ (crude in the sense that it is measured by a relatively small number of discrete categories rather than continuously). That is, we observe X, a discrete measure of θ, where $X \in \{1,2, \dots, K\}$. Let c_k be the proportion of the population with values of θ in category k or below (and where $c_0 = 0$, $c_K = 1$). Then $X = k$ if $\Phi^{-1}(c_{k-1}) < \theta \leq \Phi^{-1}(c_k)$. Note that here we have assumed no measurement error (no misclassification error).

We are interested in the relationship between some measure of student achievement, denoted Y, and θ. That is, if the relationship between Y and θ is described by the function

$Y = f(\theta) + \epsilon$, where $E(\epsilon|\theta) = 0$, we would like to estimate the function f. However, because we do not observe θ, we must infer f from the observed mean values of Y in each category of X. For example, we would like to infer the strength of the relationship between family income and test scores, given the observed mean test scores among students in each category of family income.

First, note that we can write the average value of θ within each ordinal category k as

$$\bar{\theta}_k = \frac{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} x\phi(x)dx}{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} \phi(x)dx}$$

$$= \frac{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} x\phi(x)dx}{c_k - c_{k-1}}. \tag{A1}$$

If $\phi(\theta)$ is the uniform density function, then (A1) becomes

$$\bar{\theta}_k = \frac{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} x\phi(x)dx}{c_k - c_{k-1}}$$

$$= \frac{\int_{c_{k-1}}^{c_k} xdx}{c_k - c_{k-1}}$$

$$= \frac{c_k^2 - c_{k-1}^2}{2(c_k - c_{k-1})}$$

$$= \frac{c_k + c_{k-1}}{2}. \tag{5.A2}$$

Next, note that we can write the average value of Y within each ordinal category k as

$$\bar{Y}_k = \frac{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} f(x)\phi(x)dx}{c_k - c_{k-1}}. \tag{5.A3}$$

Now suppose f can be approximated by a cubic polynomial function (one could allow a higher-order polynomial if inspection of the data suggests the need, but I find a cubic is more than sufficient for the income-achievement relationship—in many cases f is well-approximated by a line). That is,

$$Y = f(\theta) + \epsilon = a + b\theta + c\theta^2 + d\theta^3 + \epsilon, \qquad E(\epsilon|\theta) = 0.$$

(5.A4)

Now if $\phi(\theta)$ is the uniform density function, then we can express the average value of Y in category k as

$$\begin{aligned}
\bar{Y}_k &= \frac{\int_{\Phi^{-1}(c_{k-1})}^{\Phi^{-1}(c_k)} (a + bx + cx^2 + dx^3)dx}{c_k - c_{k-1}} \\
&= \frac{a(c_k - c_{k-1}) + \frac{b}{2}\left(c_k^2 - c_{k-1}^2\right) + \frac{c}{3}\left(c_k^3 - c_{k-1}^3\right) + \frac{d}{4}\left(c_k^4 - c_{k-1}^4\right)}{c_k - c_{k-1}} \\
&= a + b\frac{c_k + c_{k-1}}{2} + c\frac{c_k^2 + c_k c_{k-1} + c_{k-1}^2}{3} + d\frac{\left(c_k^2 + c_{k-1}^2\right)(c_k + c_{k-1})}{4} \\
&= a + b\bar{\theta}_k + c\frac{4\bar{\theta}_k^2 - c_k c_{k-1}}{3} + d\frac{8\bar{\theta}_k^3 - 2\left(c_k^2 c_{k-1} + c_k c_{k-1}^2\right)}{4} \\
&= a + b\bar{\theta}_k + c\frac{4\bar{\theta}_k^2 - \left(\bar{\theta}_k + \frac{c_k - c_{k-1}}{2}\right)\left(\bar{\theta}_k - \frac{c_k - c_{k-1}}{2}\right)}{3} + d\left(2\bar{\theta}_k^3 - \frac{(c_k + c_{k-1})(c_k c_{k-1})}{2}\right) \\
&= a + b\bar{\theta}_k + c\left(\bar{\theta}_k^2 + \frac{(c_k - c_{k-1})^2}{12}\right) + d\left(\bar{\theta}_k^3 + \frac{(c_k - c_{k-1})^2}{4}\right)
\end{aligned}$$

(5.A5)

We can compute the $\bar{\theta}_k$'s from (5.A2) and then can estimate a, b, c, and d by regressing the observed $\bar{Y}_k$'s on the computed $\bar{\theta}_k$'s, the $\left(\bar{\theta}_k^2 + \frac{(c_k - c_{k-1})^2}{12}\right)$'s, and the $\left(\bar{\theta}_k^3 + \frac{(c_k - c_{k-1})^2}{4}\right)$'s. The values of $\hat{a}, \hat{b}, \hat{c}$, and $\hat{d}$ describe the estimated relationship between the unobserved θ and Y. Note that these will be different, in general, than what we would get by simply regressing the $\bar{Y}_k$'s on $\bar{\theta}_k, \bar{\theta}_k^2$, and $\bar{\theta}_k^3$. The reason for the difference is that, if f is not linear, then $f(\bar{\theta}_k) \neq E[\bar{Y}_k]$.

I apply the above method to estimate the association between student test scores and two measures of parental socioeconomic status (family income and parental education).

Estimating the association between family income and achievement

In most of the studies that report income, income is reported in a set of discrete ordered categories.[2] I estimate the association between achievement and income percentile rank; this provides a comparable metric to compare income achievement gaps across time periods. Income percentile ranks have a uniform distribution, so the methods described above for uniformly distributed θ will apply. I fit cubic functions to estimate the association between achievement and income percentiles. I use weighted least squares for the estimation, weighting each observation by the inverse of the sampling variance of $\bar{Y}_k$, so that the fitted curve is influenced less by categories with small proportions of the population (and hence, large sampling variance). Below is an example of this method.

The data shown in Figure 5.A1 are from the 16-18 year-olds in the NLSY79 sample. Income is not reported categorically in NLSY; I divided reported income into 12 roughly equal sized categories here for illustration. The bars in the figure indicate the 95% confidence intervals for the mean income in a given income category; the red line represents the fitted cubic line through these data, using the method described above. Based on the fitted line, the estimated average standardized test score for a student at the 90th percentile is +0.48; the estimated average standardized test score for a student at the 10th percentile is -0.56, yielding an estimated 90/10 achievement gap of 1.04.

Figure 5.A1 here

[2] In NLSY79, NLSY97, and Add Health, parents report exact income. In NLSY97, a scatterplot of test scores against income shows a peculiar pattern: average test scores in the bottom decile are higher than in the second and third decile. Moreover, the distribution of family incomes in NLSY97 is much wider than the income distribution reported in CPS data for 1997 (in particular, there are far more families with very low income in the NLSY97 data then in the CPS data), suggesting that there may be some measurement error in the NLSY97 income data. To remedy this, I take the average family income over three years (1997-1999) for each student in the NLSY97 data in order to obtain a more reliable measure of family income. The distribution of this average income much more closely matches the CPS income distribution for 1997-1999. Moreover, the relationship between achievement and income is monotonic after this adjustment, and is similar in shape to the relationship in other samples (its first derivative is a positive, concave up function). Note that this adjustment has the effect of increasing the estimated NLSY97 income achievement gaps by roughly 10-15%, because it eliminates some attenuation bias due to measurement error, though this increase does not substantially affect the pattern of gap estimates.

More generally, I compute the 90/10 income achievement gap (the average difference in scores between a student with family income at the 90th and the 10th percentiles) as:

$$\begin{aligned}\hat{\delta}^{90/10} &= \left[\hat{Y}\middle|\theta = .9\right] - \left[\hat{Y}\middle|\theta = .1\right] \\ &= \left[\hat{a} + \hat{b}(.9) + \hat{c}(.81) + \hat{d}(.729)\right] - \left[\hat{a} + \hat{b}(.1) + \hat{c}(.01) + \hat{d}(.001)\right] \\ &= .8\hat{b} + .8\hat{c} + .728\hat{d}.\end{aligned} \tag{5.A6}$$

Likewise, the 90/50 and 50/10 income achievement gaps are:

$$\begin{aligned}\hat{\delta}^{90/50} &= \left[\hat{Y}\middle|\theta = .9\right] - \left[\hat{Y}\middle|\theta = .5\right] \\ &= \left[\hat{a} + \hat{b}(.9) + \hat{c}(.81) + \hat{d}(.729)\right] - \left[\hat{a} + \hat{b}(.5) + \hat{c}(.25) + \hat{d}(.125)\right] \\ &= .4\hat{b} + .56\hat{c} + .604\hat{d}.\end{aligned} \tag{5.A7}$$

and

$$\begin{aligned}\hat{\delta}^{50/10} &= \left[\hat{Y}\middle|\theta = .5\right] - \left[\hat{Y}\middle|\theta = .1\right] \\ &= \left[\hat{a} + \hat{b}(.5) + \hat{c}(.25) + \hat{d}(.125)\right] - \left[\hat{a} + \hat{b}(.1) + \hat{c}(.01) + \hat{d}(.001)\right] \\ &= .4\hat{b} + .24\hat{c} + .124\hat{d}.\end{aligned} \tag{5.A8}$$

I compute the standard error of each of these gap estimates from the estimated variance-covariance matrix of the regression. That is

$$\begin{aligned}se\left(\hat{\delta}^{90/10}\right) &= \left[Var\left(.8\hat{b} + .8\hat{c} + .728\hat{d}\right)\right]^{1/2} \\ &= \begin{bmatrix} 0.64Var\left(\hat{b}\right) + 0.64Var(\hat{c}) + 0.53Var\left(\hat{d}\right) \\ +1.28Cov\left(\hat{b},\hat{c}\right) + 1.165Cov\left(\hat{b},\hat{d}\right) + 1.165Cov\left(\hat{c},\hat{d}\right) \end{bmatrix}^{1/2}.\end{aligned} \tag{5.A9}$$

Because parent-reported family income is not measured perfectly, the measured income achievement gaps will differ from the true income achievement gaps by a factor of $\sqrt{r}$, where r is the reliability of the income measure. The reliability of self-reported income is typically between 0.7-1.0, with an average of 0.86 (Marquis, Marquis, & Polich, 1986), implying that the income achievement gaps will be underestimated by as much as 15%. Some newer studies suggest that family income measures based on surveys have reliabilities of .70 to .78 when compared to tax or Social Security Administration records (Coder 1992, Angrist & Kreuger, 1991). For all parent-reported income measures, I assume income is measured with reliability of 0.86.

For the studies with student-reported family income, I estimate the reliability of student-reported income measures using data from HS&B. The HS&B study includes both student-reported family income and, for a roughly 15% subsample, parent-reported family income. The existence of both student- and family-reported family income enables us to estimate the reliability of measures of income.

Let i indicate true family income. We observe a student-report (s) and a parent-report (p) of family income, each measured with error. In addition, we observe y, an error-prone measure of true student achievement, a:

$$s = i + v$$

$$p = i + u$$

$$y = a + e$$

Assuming classical measurement error in s, p, and y, the following equalities hold:

$$corr(s, p) = \sqrt{r_s \cdot r_p}$$

$$corr(s, y) = corr(i, a) \cdot \sqrt{r_s \cdot r_y}$$

$$corr(p, y) = corr(i, a) \cdot \sqrt{r_p \cdot r_y}$$

where r_s, r_p, and r_y are the reliabilities of s, p, and y, respectively. Rearranging and substituting, it follows that

$$r_s = corr(s.p) \cdot \frac{corr(s,y)}{corr(p,y)}$$

$$r_p = corr(s.p) \cdot \frac{corr(p,y)}{corr(s,y)}$$

We can observe each of these correlations, and so can use them to estimate the reliabilities of both s and p.

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. For each of these questions, respondents indicated categorical ranges of income (e.g., \$1,000-\$2,999 \$3,000-\$4,999, etc.) rather than exact dollar amounts. In order to estimate total family income, I assign a dollar amount equal to the midpoint of the category range (or \$750,000 for the top-category, which is \$500,000 or above), and then sum these amounts over all the items. Unfortunately, this results in an estimated income distribution that is much higher than the actual income distribution in 1980 (as estimated from CPS or NLSY79). For example, the 90th percentile of the income distribution computed this way is 33% higher than that reported in CPS. It may be that parents did not understand the questions well, and so double-reported some income. As a check on this, I compute a second measure of family income that is only the sum of the two wage income items (total wages of self and spouse). This yields a measure of income that is generally smaller than that computed from summing all the items, though the new measure has a much lower bottom tail (the 10th percentile of this distribution is \$425, much lower than the \$6900 10th percentile of CPS). The complexity of the parent-reported family income measure and the fact that total income must be estimated by summing the midpoints of categorical income ranges likely makes these parent-reported income measures less reliable than those in other studies, which simply ask parents their total family income.

The availability of two different parent-reported family income measures allows us to estimate the reliability of student-reported income using two different values for $corr(p, y)$. Likewise, the availability of both math and reading scores allows us to use two different outcome measures y to estimate the reliabilities of the measures. Moreover, we can estimate the reliabilities separately for 10th and 12th graders—while the reliability of the student reported measure may increase as students age, we don't expect the reliability of parent-reported income to depend on the age of the child in the sample.

The estimated reliabilities of the student- and parent-reported income measures from HS&B are reported in Table 5.A3 below. Note that the student-reported income measure has higher reliability, particularly in 12th grade, than either of the parent report measures. This is likely due to the abovementioned complexities in the HS&B parent-reported income items. Nonetheless, the reliabilities of the student-reported measures are not particularly high.

Table 5.A3 here

Note that it is also possible to estimate the reliability of student-reported income using data from twins. HS&B oversampled twins, and so contains roughly 500 pairs of twins. The reliability of student-reported income from these twin pairs is 0.69 for 10th graders and 0.75 from 12th graders, slightly higher estimates than the values I estimate above (Fetters, Stowe, & Owings, 1984). In order to be conservative, I use the lower values, (0.57 for 10th graders and 0.72 for 12th graders).

Although we cannot estimate the reliability of the parent-reported family income measures in studies other than HS&B, other published studies estimate parent-reported income measures have reliabilities of roughly 0.86, on average (Marquis, et al., 1986).

I adopt the following strategy to disattenuate the estimated gaps for measurement error in the income measure. For studies with parent-reported family income, I assume a reliability of 0.86. For studies with student reported income, I assume a reliability of 0.50 for 9th grade reports, 0.57 for 10th grade, 0.65 for 11th grade, and 0.72 for 12th grade (the 10th and 12th grade reliabilities are estimated from HS&B above; the 9th and 11th are extrapolated/interpolated from these, assuming reliability of student reports increases linearly from 9th-12th grade). For HS&B, I use the gaps estimated from student-reported family income, both because these are more reliable than the parent-reports (see above), and because this allows me to use the full HS&B sample).

To adjust for the reliability of the income measure, I multiple the computed gaps by $1/\sqrt{r}$, where r is the estimated reliability.

Because some studies measure income using many more categories than other studies (e.g., Project Talent asked students to report their family income in 5 categories; NELS asked parents to report their income in 15 categories; all other studies used at least 9 categories), we may worry that income gaps are less well estimated (and possibly biased) when income is measured using fewer categories. To check this, estimate the 90/10 income achievement gap using the NLSY79 data, in which income is reported as a continuous variable (that is, I fit a cubic model through the individual-level data to estimate the gap). I then categorize income into 20, 10, or 5 categories, and estimate the gap again using each of these categorical measures of income using the methods described above. Table 5.A4 displays the results of this exercise.

Table 5.A4 here

Table 5.A4 shows that the estimated income achievement gap does not vary much regardless of whether it is estimated using continuously- or categorically-reported income. Moreover, the estimated gap does not vary systematically with the number of categories used. Estimates based on 5, 10, or 20 categories never differ by more than one-tenth of a standard deviation from one another; nor does there appear to be any systematic direction of the differences: sometimes the gap estimates are higher when using fewer categories; sometimes they are lower. This suggests that the fact that Project Talent has income reported in only 5 categories does not systematically or substantially affect the size of the estimated gaps in Project Talent. As a result, it does not appear that differences across studies in the number of income categories used affect the estimated trend in the income achievement gap.

Estimating the association between parental education and achievement

I use two methods to estimate the association between parental education and student achievement. The first method is similar to the method used to estimate income achievement gaps. It treats the categorical measure of educational attainment as a measure of a continuous underlying latent characteristic, and then estimates the difference in average test scores between students at the 90th and 10th percentile of this distribution. This is done the same way as is described above, but uses a linear (rather than cubic) interpolation method. The second method fits a regression model predicting test scores using the categorical measure of parental education. The parental education categories reported in each study are collapsed to 4 categories (less than high school, high school diploma, some college, and BA or more) for comparability across studies. The adjusted R^2 from this regression model is used as a measure of the parental education achievement gradient.

The two approaches have different strengths; the first treats educational attainment categories as credentials. If the true distribution of human capital among parents were unchanged (other than a shift in mean human capital) and the relationship of human capital to children's achievement were also unchanged, but college degrees became more common (because of the expansion of higher education), the estimated association between parents' human capital (as measured by educational attainment) would be unchanged. That is, the linear interpolation method is unaffected by changes in credentialing rates. The R^2 method, in contrast, would be affected by changes in the proportions of parents in each credential/attainment category, because the R^2 statistic depends not only on the coefficients in the regression model but also the proportion of families in each category. On the other hand, the R^2 method treats educational attainment as a set of meaningful categories that indicate real differences in human capital (or social class). If changes in educational attainment signify real changes in the distribution of human capital, then the R^2 measure of the association between parental education and achievement would be preferred.

All 19 studies have measures of parental education, but this measure is self-reported in a number of the studies. Examination of NAEP data (in which parental education is student-reported) shows that the estimated association between parental education and achievement is much weaker for younger children; this is certainly at least partly due to the unreliability of children's report of their parents' educational attainment. As a result, I use student-reported information on parental education only if the students are in high school when they report the information.

5.A3. Estimating Income Inequality

I use CPS data to estimate the income inequality among the families of school-age children in each birth cohort. I do this as follows. Using CPS data, for each year from 1967-2008, I estimate the 90th and 10th percentiles of the income distribution of families with school-age children (ages 5-17). For these calculations, the family income distribution is weighted by the number of school-age children in a family, so the percentiles are relative to the distribution of school-age children's family incomes, rather than the distribution of family incomes (in practice, this weighting makes little appreciable difference in the estimated percentiles).

5.A4. Estimating the Trend in Income Achievement Gaps

Figures 5.1 and 5.2 in the main chapter text display the estimated achievement gaps from each of the studies with income data. Two fitted trend lines are shown in each figure: a quartic fitted line spanning the full time period; and a quadratic fitted trend using only the cohorts born from 1974-2001. Although the figures are visually useful, I also fit a set of regression models to adjust the trend line for differences among the studies.

Because the studies differ in several potentially important ways (most notably whether or not income is reported by the parents or students; the age of the cohort when the test is administered; the reliability of the tests used, and the number of income categories used in reporting income), I fit a regression model through the 90/10 income achievement gap estimates to adjust for these factors in estimating the trend in the magnitude of the gaps. The results of these models are shown in Table 5.A5.

Table 5.A5 here

The tables each show results from a set of regression models estimating the trend in the 90/10 income achievement gap based on the studies with data on family income. The regression is fit via weighted least squares, with each gap estimate weighted by the inverse of its estimated sampling variance. Standard errors are cluster-corrected for clustering of samples within studies (note that this substantially increases the size of the standard errors; without clustering, they are roughly half the size of the cluster-corrected standard errors, and the trend estimates are significant in all models except model 7).

Note that the models do not control for test reliability or for the number of categories used to measure income. Rather than control for reliability, the estimated gaps are adjusted to account for measurement error in the test (the gap estimate is multiplied by $1/\sqrt{r}$, where r is the reliability of the test). Second, because the analyses described above using NLSY data indicate that the estimated 90/10 gap is unaffected by the number of income categories used to measure income, the inclusion of a variable indicating the number of income categories used is not significant when included in the models, and so is excluded for parsimony. The gap estimates are adjusted for the estimated reliability of the family income measure, using the methods described above in section 5.A3.

Model 1 simply fits a linear trend through the data, without controls. For both reading and math, the estimated trend is roughly one-tenth of a standard deviation per decade ($p < .01$). However, this model does not control for any differences among the studies. Model 2 adds a control for whether or not family income is reported by parents or students. Because only the earliest studies—Project Talent, NLS, and HS&B—rely on student-reported income, the student-reported income dummy variable is highly correlated with the cohort variable ($r = -.72$). As a result, the inclusion of the student-report dummy variable substantially increases the standard errors. However, the student-report variable is never significant, and its inclusion does not appreciably change the trend coefficient.

Because of the collinearity of the student-report variable and the cohort variable, Model 3 excludes the three studies with student-reported family income in order to estimate the trend among those studies with parent-reported income. In both the reading and math models, the estimated trend is unchanged and remains statistically significant. Finally, Model 4 estimates the trend using only data from studies with cohorts born from 1974-2001. In these models, the trend is estimated to be roughly one-sixth of a standard deviation per decade, and is statistically different from 0 in both cases.

Models 5-8 repeat models 1-4, but add a control for the age of the students when tested to the model. In none of these models does the age coefficient approach significance; nor does it have a consistent sign across the models. Nor does its inclusion substantially alter the magnitude of the estimated trend in any of the models. However, the age variable is relatively highly correlated with the cohort variable ($r = -.61$); as a result, its inclusion increases the standard errors on the trend, so that in some cases the trend is no longer statistically significant, though the point estimates are virtually identical to the models without age included.

In sum, the regression models suggest that the 90/10 income achievement gap in reading and math increased at a significant rate from the mid 1940s to 2001, and grew particularly rapidly from 1974-2001. The estimates from Model 4 imply that the 90/10 income achievement gap grew from 0.88 in reading and 0.95 in math for the 1974 cohort to 1.27 in reading and 1.41 in math for the cohort born in 2001, an increase of 40-50% over less than 3 decades.

Another way of examining the trend in the income achievement gap is to look at the gap as measured in studies with similar tests. There are six studies conducted by the National Center for Education Statistics in our sample (NLS, HS&B, NELS, ELS, ECLS-K, and ECLS-B), spanning birth cohorts from 1954 to 2001. The tests used in these studies are similar to one another, using many overlapping items.[3] Table 5.A6 shows the estimated 90/10 income achievement gaps from each of these NCES studies. The estimated gaps are relatively constant in size from the NLS through NELS cohorts (born 1954-1974), grow somewhat by the ELS cohort (born 1986) and then grow substantially by the two ECLS cohorts (born 1992-20019, increasing from roughly 0.9-1.0 standard

---

[3] The reading tests given in NLS and to the HS&B 1980 seniors were identical. Eighteen of 25 items that appeared in the NLS math test also appeared on the HS&B math test given to 1980s seniors. Eight of 19 items on the reading test given in NLS and to HS&B 1980 seniors were also given to HS&B 1982 seniors. Seventeen of 28 items on the math test given to HS&B 1982 seniors were also given to HS&B 1980 seniors (Rock, et al., 1985).

Some of the NELS items overlap with prior assessments, including those administered in HS&B, NAEP, SIMS, ETS test files from previous operational tests, and a pool of items written specifically for the NELS: 88 battery. The NELS math test contained 16 items in common with the HS&B test (out of 81 items on NELS) (Rock, Pollack, & Quinn, 1992). Some of the ELS test questions were selected from previous assessments, including those administered in NELS, NAEP, and PISA (Ingels, et al., 2005). The NELS and ELS math tests shared 44 common items.

The ECLS-K assessments included items that were specifically created for the study, items adapted from commercial assessments with copyright permission, and other NCES studies including items from NAEP (disclosed items), NELS, and ELS (Najarian, Pollack, Sorongon, & Hausken, 2009).

deviations in the NLS, HS&B, and NELS cohorts to 1.2-1.4 standard deviations in the ECLS-K and ECLS-B cohorts.

Table 5.A6 here

The other possible comparison of samples given similar tests are the NLSY79 and NLSY97 cohorts (born in the early 1960s and early 1980s). The estimated income achievement gaps in the two NLSY studies are essentially unchanged across this 20-year period (see Table 5.A6). This result is consistent with the trend in the NCES studies, at least through the mid 1970s cohorts, The NCES studies do show a rise in the gap between 1974 and the 1986 cohorts, while the NLSY studies show no rise through the early 1980s. It may be that most of the observed rise in the NCES studies occurs in the 1980s, which would not be observed in the NLSY studies. Or it may be that the content of the AFQT (the test used in the NLSY studies) is less responsive to the income or income-related factors that drive the upward trend in the NCES (and other studies). The NCES tests are designed to be tests of academic achievement—they deliberately measures content that is taught in schools. The AFQT, on the other hand, has been described by some as a measure of ability (latent cognitive skill, or fluid intelligence). Several studies, however, have demonstrated that AFQT scores are affected by schooling, and so at least partly are measures of achievement (Cascio & Lewis, 2005; Cordero-Guzman, 2001; Hansen, Heckman, & Mullen, 2004; Roberts, et al., 2000). A closer analysis of the differences in the content of the AFQT and the NCES tests might shed more light on whether the differences in trends in the income achievement gaps are partly due to differences in the content of the tests.

A final concern is the fact that some of the studies (Talent, NLS, HS&B, Add Health, and ELS) include school-based samples of high school students. Because some students dropout (and those who dropout have lower levels of academic achievement and come from lower-income families, on average, than those who complete high school), the exclusion of dropout from the early studies may bias the estimated achievement gaps downward. Moreover, because most of these studies are from the earlier cohorts, when dropout rates were higher, this may bias the estimated trend in the gaps. The NELS data, however, provide a method of testing how severe such bias may be. NELS sampled eighth-grade students and tested them in math and reading in 8th, 10th, and 12th grade. Because few students dropout before the end of eighth grade, the full NELS sample is representative of its age cohort. Although some students in the NELS sample dropped out of school before the end of high school, a random subsample of these students was followed-up, surveyed, and tested in math and reading in the years when they would have been in 10th and 12th grade. Thus, we can compare the estimated income achievement gap in 12th grade based on the both the full NELS sample (by using the appropriate weights to adjust for the probability sampling of dropouts) and the NELS sample that was still in school in 12th grade. The latter mimics the type of sample we observe in Talent, NLS, HS&B, Add Health, and ELS (where we only observe test scores for students still in school); the former is what we would like to be able to estimate (the gap in the full cohort population). Likewise, we can compute the gaps in 8th grade from the full cohort sample and only among those who persisted in school through 12th grade. In both cases, the difference between the full cohort and persister estimates provides some guidance regarding the likely size of the bias in the early studies that do not include high school dropouts.

Table 5.A7 here

The results of these exercises are shown in Table 5.A7. I compute the 90/10, 90/50, and 50/10 income achievement gaps as well as White-Black and White-Hispanic gaps. In no case do the estimated gaps in the full cohort and among the persister sample differ by more than .02 standard deviations. This suggests that there is no meaningful bias introduced into the school-based high school samples (Talent, NLS, HS&B, Add Health, and ELS).

Trends by race and gender

For most of the studies, it is possible to estimate the 90/10 income gap separately among the white, black, and Hispanic student populations, and separately by student gender. These are shown in Figures 5.A2-5.A11. Fitted trends, weighted by the inverse of the sampling variance of the estimated gaps, are shown. Note that the 90/10 gaps here are between students at the 90th and 10th percentiles of the income distribution of the full population, not the 90th and 10th percentiles of the specific (race or gender) group's family income distribution. Thus, the gaps here are comparable to those in the full population.

Figures 5.A2-5.A11 here

5.A5. Comparing income gaps to race gaps

The estimated race gap trends from the 12 studies with income shown in Figures 5.3 and 5.4 is a quadratic fitted trend line. The estimated race gap trend from the NAEP studies is a polynomial fitted trend line (quartic for reading; cubic for math, because the quartic term is not significant in the math models), adjusted for the age of the students when tested and controlling for whether the gaps come from the Long-Term Trend NAEP or Main NAEP. The fitted lines show the estimated black-white gaps for 13-year-olds on the LTT-NAEP tests.

The black-white gap widens considerably in the early elementary grades, which explains why the black-white gap in the ECLS-K and ECLS-B is so much smaller than the contemporaneous NAEP gap (which is based on gaps at age 9 and 13).

5.A6. Does rising inequality account for the growth of the income achievement gap?

Income inequality grew substantially in the U.S from the 1970s to the present. Figure 5.A12 displays the long-term trend in income inequality in the U.S., as measured by the share of total income accruing to the top 10% of earners. Note that inequality was relative low and stable from the mid 1940s into the 1970s, when it began to rise rapidly.

Figure 5.A12 here

To formally test whether income inequality trends account for the income achievement gap trends, I fit a series of regression models of the form

$$\delta_i^{90/10} = \gamma_0 + \gamma_1(COH_i) + \gamma_2\left(\log_2\left(\frac{I_i^{90}}{I_i^{10}}\right)\right) + e_i \tag{5.A10}$$

where i indexes study cohorts, δ_i is the estimated 90/10 income achievement gap, COH_i is an indicator for the birth year of a given study cohort (centered on 1974), and where $R_i^{90/10}$ is the 90/10 income ratio of study cohort i. Rather than estimate the 90/10 income ratio from the specific sample, I use CPS data to estimate the 90/10 ratio. Because it is not immediately obvious whether it is income inequality in recent years, in the early years of one's life, or over the course of one's entire life that matter the most, I use different versions of the 90/10 ratio in the models. In some models $R^{90/10}$ is the 90/10 income ratio in the year the test is given; in some it is the average of the 90/10 ratios for the 5 prior years; in some it is the average 90/10 ratio for the first 5 years of a child's life (e.g., for the NELS sample—14 years olds born in 1974 and tested in 1988—this would be the average of the 90/10 ratio in 1974-1978); and in some I use the average of the 90/10 over the child's entire life. Because CPS income data are available only from 1967 forward, I do not include in these models cohorts born prior to 1963 (I exclude the NLS and Project Talent samples). For HS&B and NLSY79—cohorts born in the early/mid 1960s, I use the data for the years available.

I fit a second set of models, identical to those described above, but using the 90/50 and 50/10 income achievement gaps and income ratios in place of the 90/10 gaps and ratios. These models test whether the growing 50/10 income inequality accounts for the growing 50/10 income gaps (and likewise for the 90/50 models).

In general, these models suggest that income inequality does not explain the rising income achievement gap. The estimates are shown in Tables 5.A8 and 5.A9 below. In general, the coefficients on the logged income ratio are very unstable across the models, but also have large standard errors and so are significant in only one of 24 models. The income inequality measures are highly correlated ($r > 0.9$) with the cohort birth year, making the estimated coefficients very imprecise due to multicollinearity. Nonetheless, the estimated trend in the 90/10 income achievement gap is as large or larger as in the models shown above that do not control for income inequality; the trend estimates range from 0.10 to 0.22 standard deviations per decade, net of income inequality trends. In reading, these 90/10 trend coefficients are statistically significant ($p < .10$) in each of models; in the math models they are never significant at $p < .05$ in only two of the models, but are of roughly the same magnitude as the reading coefficients.

Tables 5.A8 and 5.A9 here

The bottom panels of Tables 5.A8 and 5.A9 show there is little or no evident trend in the 50/10 income achievement gap over this time period, particularly in reading. There is, however, evidence that the 90/50 income achievement gap has grown, particularly in reading (middle panel, Table 5.A7). Growing income inequality explains none of that growth; in fact the estimated trends are steeper, controlling for income inequality.

5.A7. Does an Increasing Association Between Income and Achievement Explain the Growth of the Income Achievement Gap?

In this section I estimate the association between family income (in logged dollars) and child achievement. Although I refer to this association in some places as an estimate of the "achievement returns to income," this parameter should not be thought of as a causal parameter. I estimate the association between income and achievement indirectly. Suppose the relationship between income (I) and achievement (A) is given by

$$A = \beta_0 + \beta_1 \log_2(I) + e.$$

(5.A11)

Then the 90/10 income achievement gap can be written

$$\begin{aligned}\delta^{90/10} &= E[\bar{A}|I = I^{90}] - E[\bar{A}|I = I^{10}] \\ &= \beta_1 \log_2\left(\frac{I^{90}}{I^{10}}\right) \\ &= \beta_1 \log_2\left(R^{90/10}\right),\end{aligned}$$

(5.A12)

where I^{90} and I^{10} are the incomes at the 90th and 10th percentiles of the income distribution, respectively. Thus, the gap is a function of both the strength of the income coefficient (β_1) and the extent of income inequality ($R^{90/10} = I^{90}/I^{10}$). An increase in the achievement gap, therefore, might result from an increase in inequality, an increase in the strength of the association between income and achievement, or some combination of the two. Rather than estimate β_1 from a regression of individual-level achievement on logged income, I note that

$$\beta_1 = \frac{\delta^{90/10}}{\log_2(R^{90/10})}.$$

(5.A13)

I use the estimated 90/10 income gap from each study, and divide it by the logged 90/10 income ratio (computed from CPS data, as described above) to estimate β_1. This has two advantages over using individual-level data and regressing achievement on income. First, it allows me to estimate the gap using a nonlinear (cubic) model, and it yields less noisy estimates of β_1 because the CPS income inequality estimates are much more precise than I obtain from each individual study.

I compute the β_1 for each of the samples in this way. I also compute the corresponding coefficient for the 50/10 and 90/50 regions of the income distribution. For example,

$$\hat{\beta}_1^{50/10} = \frac{\hat{\delta}^{50/10}}{\log_2\left(\hat{R}^{50/10}\right)}$$

(5.A14)

is the estimated association between income and achievement for families with incomes below the median income. Likewise,

$$\hat{\beta}_1^{90/50} = \frac{\hat{\delta}^{90/50}}{\log_2\left(\hat{R}^{90/50}\right)}$$

(5.A15)

is the estimated association between income and achievement for families with incomes above the median. Figures 5.A13-5.A18 below display the estimated associations between income and achievement implied by the equations above. In these figures, I use the average income inequality during the five years prior to the test year,[4] though the figures are very similar regardless of which income ratio I use. The figures show very little change in the income-achievement association when we consider the full income distribution (Figures 5.A13 and 5.A14). However, when we consider the association among families with incomes above the median, it appears that the "returns to income" have grown considerably over the last several decades, particularly in reading, where the coefficient has increased by 50-60% in the last 25 years (Figures 5.A15 and 5.A16). This is not true when we examine the trend in the estimated association between income and achievement for families below the median income (Figures 5.A17 and 5.A18). Among these families, the "returns to income" have been the same—or even declining—for 50 years.

Figures 5.A13-5.A18 here

5.A8. Estimating Trends in Parental Education Gaps

Figures 5.A19 and 5.A20 display the association between parental educational attainment and math and reading scores from all available studies. Studies where parental educational attainment is student-reported are included only if the students were in high school at the time of

[4] For Project Talent I use the 90/10 ratio estimated from the sample, as there is no CPS data for this cohort. For NLS, I use the 90/10 ratio from the CPS averaged over the years 1967-1972, the only years for which this cohort has CPS data.

the report (this means, for example, that only age 17 NAEP-LTT and grade 12 Main NAEP estimates are included here). A reliability of 0.90 is assumed for high-school student reports of parental educational attainment, consistent with twin-based estimates obtained from the HS&B data (Fetters, et al., 1984) Each figure includes three fitted trend lines: i) a trend line based on the high school NAEP data; ii) a trend line based on the estimated parental education gap from 16 non-NAEP studies; and iii) a fitted cubic trend line based on all estimates, but adjusted for the age of the students when tested and whether parental education was reported by students or parents. Each of these fitted trends suggests a slightly different story.

Figures 5.A19 and 5.A20 here

NAEP data suggest that the parental education gap changed little or declined slightly across cohorts born from 1960 to 1990. The non-NAEP studies, in contrast, suggest that the parental education achievement gap grew by roughly 30% among cohorts born from the 1940s to 2001. This trend, however, may be confounded by the fact that early studies relied on student-reported parental education and that later studies had a wider age range of students. The adjusted trend line is based on a regression model that controls the age at which students were tested, and for whether or not parental education was student- or parent-reported. The fitted line displays the estimated trend in the parental education gap for 14-year-old students in studies with parent-reported educational attainment. For both math and reading, these trends show little change from the 1940s through the 1990s; while there is some suggestion that the parental education gap may have increased among the most recent cohorts, there are too few data points in the last decade to be sure of this.

On the whole, the data suggest that the association between parental educational attainment and student achievement has not changed dramatically over the last 50 years, though there is some evidence that it may be increasing in recent decades.

To estimate the partial associations between income, parental education, and achievement, I fit regression models of the form

$$Y_i = \beta_0 + \beta_1(INC_i) + \beta_2(PARED_i) + \mathbf{R}_i\mathbf{\Gamma} + \epsilon_i \tag{5.A16}$$

where INC_i is family income measured in percentiles; $PARED_i$ is parental educational attainment measured in percentiles;[5] and $\mathbf{R}_i$ is a vector of race dummy variables. I multiple both $\hat{\beta}_1$ and $\hat{\beta}_2$ by 0.8 so that they can be interpreted as the average difference in achievement between students at the 90th and 10th percentiles of the income (or parental education) distributions, controlling for parental education (or income) and race. This makes them comparable to the income gaps reported in Figures 5.1 and 5.2. I then plot the estimated coefficients $0.8 \cdot \hat{\beta}_1$ and $0.8 \cdot \hat{\beta}_2$ from each study and wave across cohorts in Figures 5.10 and 5.11.

[5] I measure both income and parental education in percentiles so that the coefficients can be compared to one another. To measure income in percentiles, I assign each student the income percentile corresponding to the middle percentile of the income category of his family income (for example, if 30% of students are in income categories $1, 2, \ldots k-1$, and if 10% of students are income category k, then all students in income category k are assigned income percentile of $\left(0.30 + \frac{1}{2} \cdot 0.10\right) = 0.35$). To measure parental education in percentiles, I assign each student the maximum of his or her father's and mother's educational attainment category, and convert these ordered categories into percentiles in the same was as I do for income categories.

References

Cascio, E., & Lewis, E. (2005). Schooling and the AFQT: Evidence from school entry laws. *Federal Reserve Bank of Philadelphia Working Paper, 05-1*.

Cordero-Guzman, H. (2001). Cognitive skills, test scores, and social stratification: The role of family and school-level resources on racial/ethnic differences in scores on standardized tests. *The Review of Black Political Economy, 28*(4), 31-71.

Fetters, W. B., Stowe, P. S., & Owings, J. A. (1984). *Quality of Responses of High School Students to Questionnaire Items*. Washington, DC: National Center for Education Statistics.

Hansen, K., Heckman, J., & Mullen, K. (2004). The effect of schooling on ability and achievement test scores. *Journal of Econometrics, 121*(1-2), 39-98.

Ingels, S. J., Owings, J. A., Pratt, D. J., Rogers, J. E., Siegel, P. H., & Stutts, E. S. (2005). *Education Longitudinal Study of 2002: Base-Year to First Follow-Up Data File Documentation*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Jencks, C., Owens, A., Shollenberger, T., & Zhu, Q. (2010). *How Has Rising Economic Inequality Affected Children's Educational Outcomes?*

Marquis, K. H., Marquis, M. S., & Polich, J. M. (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association, 81*(394), 381-389.

Najarian, M., Pollack, J. M., Sorongon, A. G., & Hausken, E. G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: Psychometric Report for Eighth Grade*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Roberts, R., G., G., Anjoui, F., Kyllonen, P., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Battery (ASVAB): Little more than acculturated learning? . *Learning and Individual Differences, 12*(1), 81-103.

Rock, D. A., Hilton, T. I., Pollack, J. M., Ekstrom, R. B., Goertz, M. E., & Maw, C. (1985). *Psychometric Analysis of the NLS and the High School and Beyond Test Batteries*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Rock, D. A., Pollack, J. M., & Quinn, P. (1992). *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up*. Washington, DC: National Center for Education Statistics, U.S. Department of Education Office of Educational Research and Improvement.

APPENDIX TABLES & FIGURES

Table 5.A1: Studies Used

Name of Study	Abbreviation	Year(s) Tested	Age(s) Tested	Grade(s) Tested	Test/ Subject(s)	Sample Size	Income Data Available
Early Childhood Longitudinal Study-Birth Cohort	ECLS-B	2001-2007	1-6	pre-k/k	reading, math	10,700	Yes
Early Childhood Longitudinal Survey-Kindergarten	ECLS-K	1998-2007	6-11	k-8	reading, math	24,500	Yes
Education Longitudinal Study	ELS	2002, 2004	16, 18	10, 12	reading, math	15,300	Yes
Equality of Educational Opportunity	EEO	1966	7,9,12,15,18	1,3,6,9,12	reading, math	76,000-134,000	No
High School and Beyond	HS&B	1980, 1982	16, 18	10, 12	reading, math	30,000	Yes
High School and Beyond	HS&B	1980	18	12	reading, math	28,000	Yes
Longitudinal Survey of American Youth	LSAY	1987-1992	13-18	7-12	math	3,100	No
Longitudinal Survey of American Youth	LSAY	1987-1990	16-18	10-12	math	2,800	No
NAEP-Long Term Trend	NAEP-LTT	1971-2004	9, 13, 17	3, 7, 11	reading, math	4,000-25,000	No
NAEP- MAIN	Main NAEP	1990-2007	10, 14, 18	4, 8, 12	reading, math	8,000-180,000	No
National Education Longitudinal Study	NELS	1988, 1990, 1992	14, 16, 18	8, 10, 12	reading, math	20,000-25,000	Yes
National Longitudinal Study	NLS	1972	18	12	reading, math	16,683	Yes
National Longitudinal Study of Adolescent Health	Add Health	1995	13-18	7-12	PPVT (vocabulary)	21,000	Yes
National Longitudinal Survey of Youth: 79	NLSY79	1981	16-18	10th-12th	ASVAB	12,000	Yes
National Longitudinal Survey of Youth: 97	NLSY97	1998	12-18	6-11	ASVAB	9,000	Yes
Program of International Student Assessment	PISA	2000, 2003, 2006	15	10	reading, math	3,800-5,600	No
Progress in International Reading Study	PIRLS	2001, 2006	9	4	reading	3,600-5,200	No
Project Talent	Talent	1960	14-18	9-12	reading, math	377,000	Yes
Prospects	Prospects	1991-1994	6-15	1-9	reading, math	12,000-20,000	Yes
Study of Early Child Care and Youth Development	SECCYD	1994-2006	3-15	pre-k - 9	Bayley, BBCS, WJ	1,000-1,300	Yes
Third International Mathematics and Science Study	TIMSS	1995	9, 13, 17	3, 8, 12	math	33,000	No
Third International Mathematics and Science Study-Repeat	TIMSS	1999	13	8	math	9,072	No
Trends in International Math and Science	TIMSS	2003, 2007	9, 13	4, 8	math	15,000-19,000	No

PPVT: Peabody Picture Vocabulary Test; ASVAB: Armed Services Vocational Aptitude Battery; BBCS: Bracken Basic Concept Scale; WJ: Woodcock Johnson

Note: family income is reported by students in Project Talent, NLS, and HS&B.

Table 5.A2

Test Reliabilities in Math and Reading

	Math	Reading	Type	Source
Early Childhood Longitudinal Study-Birth Cohort (Pre-K)	0.89	0.84	IRT	http://nces.ed.gov/pubs2010/2010009.pdf
Early Childhood Longitudinal Study-Kindergarten (Fall-K)	0.91	0.92	IRT	User's Manual for the ECLS-K Eighth-Grade
Education Longitudinal Study (2002)	0.92	0.86	IRT	Ingels et al. 2006
Equality of Educational Opportunity Study	0.75	0.75	---	See note
High School & Beyond: 1980 Seniors	0.85	0.79	IC	Rock et al. 1985: Page 47-49
High School & Beyond: 1980 Sophomores	0.87	0.77	IC	Rock et al. 1985: Page 47-49
Longitudinal Survey of American Youth	0.96	NA	---	See note
NAEP-Long Term Trend (1996)	.85-.93	NF	IRT	http://nces.ed.gov/nationsreportcard/pdf/main1996/1999452d.pdf
NAEP-Long Term Trend (2003)	NF	.82-.86	IC	http://nces.ed.gov/nationsreportcard/pdf/main1998/2001509c.pdf
NAEP- MAIN (1996)	.95-.98	NF	IRT	http://nces.ed.gov/nationsreportcard/pdf/main1996/1999452d.pdf
NAEP-MAIN (1998)	NF	.70-.74	IC	http://nces.ed.gov/nationsreportcard/pdf/main1998/2001509c.pdf
NAEP- MAIN (2003)	NF	.72-.76	IC	http://nces.ed.gov/nationsreportcard/tdw/analysis/initial_classical.asp#table1
National Education Longitudinal Study (1988)	0.90	0.84	IC	Rock & Pollack 1991
National Longitudinal Study	0.86	0.79	IC	Rock et al. 1985: Page 47-49
National Longitudinal Study of Adolescent Health (PPVT)	NA	0.95	IC	
National Longitudinal Survey of Youth: 79	.84-.86	.75-.88	IRT	http://officialasvab.com/reliability_res.htm#table3
National Longitudinal Survey of Youth: 97	0.93	.86-.93	IRT	http://officialasvab.com/reliability_res.htm#table3
Program of International Student Assessment (2000)	0.81	0.89	IC	http://www.oecd.org/dataoecd/53/19/33688233.pdf
Program of International Student Assessment (2003)	0.85	0.80	IC	http://www.pisa.oecd.org/dataoecd/49/60/35188570.pdf
Program of International Student Assessment (2006)	0.78	0.78	IC	http://www.oecd.org/dataoecd/0/47/42025182.pdf
Progress in International Reading Study (2001)	NA	0.9	IC	Mullis et al. 2003, p. 298
Progress in International Reading Study (2006)	NA	0.88	IC	Mullis et al. 2007, p. 306
Project Talent	0.75	0.75	---	See note
Prospects (3rd Grade)	0.8	0.85	TR	http://epm.sagepub.com/cgi/reprint/61/5/841.pdf
Study of Early Child Care and Youth Development (WJ)	.86-.95	.88-.94	IC	http://www.iapsych.com/wj3ewok/LinkedDocuments/asb-2.pdf
Third International Mathematics and Science Study (2003, 4th Grad	0.88	NA	IC	Mullis et al. 2004, p. 368
Third International Mathematics and Science Study (2003, 8th Grad	0.9	NA	IC	Mullis et al. 2004, p. 368)
Third International Mathematics and Science Study (1999, 8th Grad	0.9	NA	IC	Mullis et al. 2000, p. 333
Third International Mathematics and Science Study (1995, 3rd Grad	0.83	NA	IC	Mullis et al. 1997, p. A-24
Third International Mathematics and Science Study (1995, 4th Grad	0.86	NA	IC	Mullis et al. 1997, p. A-24
Third International Mathematics and Science Study (1995, 7th Grad	0.89	NA	IC	Beaton et al. 1996, p. A-26
Third International Mathematics and Science Study (1995, 8th Grad	0.89	NA	IC	Beaton et al. 1996, p. A-26
Third International Mathematics and Science Study (1995, 12th Gra	0.8	NA	IC	Mullis et al. 1998, p. B-37

Note: NF = Not found; NA=not applicable; IC = internal consistency; TR = test-retest; IRT = reliability of IRT estimate. I was unable to locate reliabilities for all administrations of Main and Long-Term Trend NAEP. The ASVAB reliabilities were computed for 2005 military applicants as I was unable to locate reliabilities for the NLSY sample members. The reliabilities listed for NLSY 1979 refer to reliabilities for paper and pencil administrations of the test and the reliabilities listed for NLSY 1997 refer to reliabilities for comptuer adaptive administrations of the test. I was unable to find information on the reliability of the tests in Project Talent and EEO; for both I assumed a reliability of 0.75, which is slightly lower than the lowest-reported reliability of any of the other studies. I was unable to find reliabilities for the Longitudinal Survey of American Youth; this study's test was made up of NAEP items so I use a reliability of 0.96 which is the average (across grades) of the reliabilities of the 1996 administration of the math portion of the Main-NAEP.

Table 5.A3

Estimated reliabilities of student- and parent-reported family income measures, HS&B

	using all-item parent report		using wage only parent report		
income measure	**math**	**reading**	**math**	**reading**	**average**
student-report, 10th grade	0.57	0.49	0.70	0.53	0.57
student-report, 12th grade	0.67	0.65	0.79	0.78	0.72

	using 10th grade		using 12th grade		
income measure	**math**	**reading**	**math**	**reading**	**average**
parent-report, all items	0.40	0.47	0.48	0.50	0.46
parent-report, wages only	0.31	0.42	0.35	0.36	0.36

Table 5.A4

Estimated 90/10 Income Achievement Gap, NLSY79, Using Different Categorizations of Income

	Reading			Math		
Income Measure	Age 16	Age 17	Age 18	Age 16	Age 17	Age 18
Continuous	0.995	1.079	1.040	1.080	1.176	1.223
	(0.114)	(0.080)	(0.082)	(0.110)	(0.086)	(0.082)
20 Categories	0.985	1.094	1.069	1.048	1.172	1.257
	(0.110)	(0.072)	(0.108)	(0.110)	(0.078)	(0.072)
10 Categories	0.982	1.065	1.045	1.039	1.158	1.231
	(0.100)	(0.050)	(0.096)	(0.096)	(0.062)	(0.076)
5 Categories	0.951	1.126	1.134	1.053	1.184	1.304
	(0.068)	(0.025)	(0.100)	(0.111)	(0.132)	(0.137)

Table 5.A5

Estimated Trend in 90/10 Income Gap, Reading

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Cohort Birth Year	0.010 **	0.011 *	0.010 +	0.015 **	0.010 **	0.011 *	0.010	0.017 **
	(0.001)	(0.004)	(0.005)	(0.003)	(0.001)	(0.005)	(0.007)	(0.004)
Student-Reported Income		0.032				0.048		
		(0.180)				(0.207)		
Age at Test					0.000	0.002	0.001	0.007
					(0.007)	(0.008)	(0.012)	(0.007)
Intercept	0.951 **	0.941 **	0.951 **	0.882 **	0.951 **	0.939 **	0.950 **	0.866 **
	(0.024)	(0.065)	(0.072)	(0.036)	(0.025)	(0.072)	(0.089)	(0.039)
Studies Included								
TALENT, NLS, HS&B	X	X			X	X		
NLSY79	X	X	X		X	X	X	
Post-1970 studies	X	X	X	X	X	X	X	X
N (samples)	26	26	19	17	26	26	19	17
N (studies)	12	12	9	8	12	12	9	8
R-squared	0.911	0.912	0.396	0.720	0.911	0.912	0.396	0.736

Note: Standard errors in parentheses, corrected for clustering of samples within studies. + p<.10; * p<.05; ** p<.01. Cohort birth year is centered at 1974; age is centered at 13, so intercept describes estimated 90/10 income achievment gap among 13-year-olds born in 1974. Observations are weighted by the inverse of their squared standard errors.

Estimated Trend in 90/10 Income Gap, Math

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Cohort Birth Year	0.009 **	0.012 *	0.013 +	0.017 *	0.008 **	0.012 *	0.012	0.017
	(0.002)	(0.005)	(0.006)	(0.007)	(0.001)	(0.005)	(0.009)	(0.009)
Student-Reported Income		0.161				0.157		
		(0.205)				(0.216)		
Age at Test					-0.004	-0.001	-0.003	0.000
					(0.012)	(0.012)	(0.018)	(0.016)
Intercept	1.075 **	1.024 **	1.017 **	0.949 **	1.070 **	1.024 **	1.020 **	0.948 **
	(0.047)	(0.084)	(0.097)	(0.099)	(0.039)	(0.087)	(0.117)	(0.116)
Studies Included								
TALENT, NLS, HS&B	X	X			X	X		
NLSY79	X	X	X		X	X	X	
Post-1970 studies	X	X	X	X	X	X	X	X
N (samples)	20	20	13	11	20	20	13	11
N (studies)	11	11	8	7	11	11	8	7
R-squared	0.808	0.819	0.372	0.542	0.810	0.819	0.373	0.542

Note: Standard errors in parentheses, corrected for clustering of samples within studies. + p<.10; * p<.05; ** p<.01. Cohort birth year is centered at 1974; age is centered at 13, so intercept describes estimated 90/10 income achievment gap among 13-year-olds born in 1974. Observations are weighted by the inverse of their squared standard errors.

Table 5.A6

90/10 Income achievement gaps, by study and subject, all NCES studies and NLSY samples

Study Sample	Birth Year	Test Year	Reading	Math
NLS	1954	1972	0.919	1.006
			(0.060)	(0.043)
HS&B Grade 12	1962	1980	0.783	0.912
			(0.093)	(0.075)
HS&B Grade 10	1964	1980	0.938	1.023
			(0.102)	(0.102)
NELS	1974	1988	0.885	1.023
			(0.029)	(0.032)
ELS	1985	2002	1.094	1.111
			(0.039)	(0.041)
ECLSK	1993	1998	1.229	1.397
			(0.028)	(0.025)
ECLSB	2001	2006	1.198	1.280
			(0.037)	(0.045)
NLSY79	1963-65	1981	1.205	1.266
			(0.058)	(0.073)
NLSY97	1981-85	1997	1.198	1.241
			(0.040)	(0.048)

Table 5.A7

Achievement Gaps Among All Full Cohort (Including Dropouts) and Among Those Who Persist to 12th Grade (NELS)

	Income Achievement Gap			Race Achievement Gap	
	90/10	90/50	50/10	White-Black	White-Hispanic
Grade 12 Achievement Gaps					
Math					
Full Cohort Gap Estimate	1.05	0.63	0.42	0.86	0.61
Estimated Gap Among Students Who Persist to Grade 12	1.05	0.61	0.43	0.86	0.61
Reading					
Full Cohort Gap Estimate	0.91	0.48	0.42	0.73	0.60
Estimated Gap Among Students Who Persist to Grade 12	0.89	0.46	0.45	0.73	0.60
Grade 8 Achievement Gaps					
Math					
Full Cohort Gap Estimate	1.02	0.53	0.49	0.86	0.66
Estimated Gap Among Students Who Persist to Grade 12	1.02	0.51	0.51	0.90	0.66
Reading					
Full Cohort Gap Estimate	0.88	0.39	0.49	0.76	0.62
Estimated Gap Among Students Who Persist to Grade 12	0.87	0.38	0.51	0.79	0.62

Note: gaps are disattenuated to correct for measurement error in test or measurement error in reported family

Table 5.A8

Estimated Association Between Income Inequality and the Income Achievement Gap, Reading

		Measure of Income Inequality Used			
	Base Model	Current Year 90/10 Ratio	5 Year Average 90/10 Ratio	Average 90/10 Ratio When Ages 0-4	Average 90/10 Ratio Over Whole Life
90/10 Income Achievement Gap					
Cohort Birth Year	0.010 *	0.019 **	0.015 **	0.015 +	0.018 *
	(0.004)	(0.006)	(0.004)	(0.007)	(0.007)
$\log_2$(90/10 Ratio)		-1.339	-0.665	-0.202	-0.483
		(0.756)	(0.399)	(0.332)	(0.468)
R-squared	0.424	0.633	0.556	0.451	0.506
90/50 Income Achievement Gap					
Cohort Birth Year	0.011 **	0.011 +	0.015 +	0.031 +	0.028 *
	(0.002)	(0.005)	(0.007)	(0.015)	(0.009)
$\log_2$(90/50 Ratio)		-0.012	-0.562	-1.754	-1.946
		(0.824)	(1.268)	(1.277)	(1.165)
	0.573	0.579	0.590	0.634	0.642
50/10 Income Achievement Gap					
Cohort Birth Year	-0.001	-0.001	-0.001	-0.003 +	-0.002
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
$\log_2$(50/10 Ratio)		-0.214	-0.251	0.164	0.072
		(0.423)	(0.326)	(0.241)	(0.368)
	0.010	0.050	0.074	0.076	0.033
N (samples)	21	21	21	21	21
N (studies)	10	10	10	10	10

Note: Standard errors in parentheses, corrected for clustering of samples within studies. + p<.10; * p<.05; ** p<.01. Project Talent and NLS studies excluded for lack of income inequality data. Observations are weighted by the inverse of their squared standard errors.

Table 5.A9

Estimated Association Between Income Inequality and the Income Achievement Gap, Math

			Measure of Income Inequality Used							
	Base Model		Current Year 90/10 Ratio		5 Year Average 90/10 Ratio		Average 90/10 Ratio When Ages 0-4		Average 90/10 Ratio Over Whole Life	
90/10 Income Achievement Gap										
Cohort Birth Year	0.012	*	0.018	+	0.016	+	0.022		0.022	
	(0.005)		(0.009)		(0.008)		(0.015)		(0.016)	
$\log_2$(90/10 Ratio)			-0.800		-0.327		-0.342		-0.551	
			(1.003)		(0.803)		(0.451)		(0.738)	
R-squared	0.402		0.459		0.420		0.449		0.454	
90/50 Income Achievement Gap										
Cohort Birth Year	0.007	+	0.004		0.011		0.038	**	0.032	
	(0.003)		(0.008)		(0.009)		(0.009)		(0.020)	
$\log_2$(90/50 Ratio)			0.536		-0.577		-2.671	**	-2.782	
			(0.872)		(1.289)		(0.616)		(2.226)	
	0.257		0.278		0.268		0.366		0.352	
50/10 Income Achievement Gap										
Cohort Birth Year	0.004		0.005		0.005		0.003		0.003	
	(0.003)		(0.004)		(0.005)		(0.007)		(0.007)	
$\log_2$(50/10 Ratio)			-0.303		0.049		0.147		0.176	
			(0.628)		(0.686)		(0.366)		(0.613)	
	0.143		0.173		0.149		0.166		0.159	
N (samples)	15		15		15		15		15	
N (studies)	9		9		9		9		9	

Note: Standard errors in parentheses, corrected for clustering of samples within studies. + p<.10; * p<.05; ** p<.01. Project Talent and NLS studies excluded for lack of income inequality data. Observations are weighted by the inverse of their squared standard errors.

Figure 5.A1

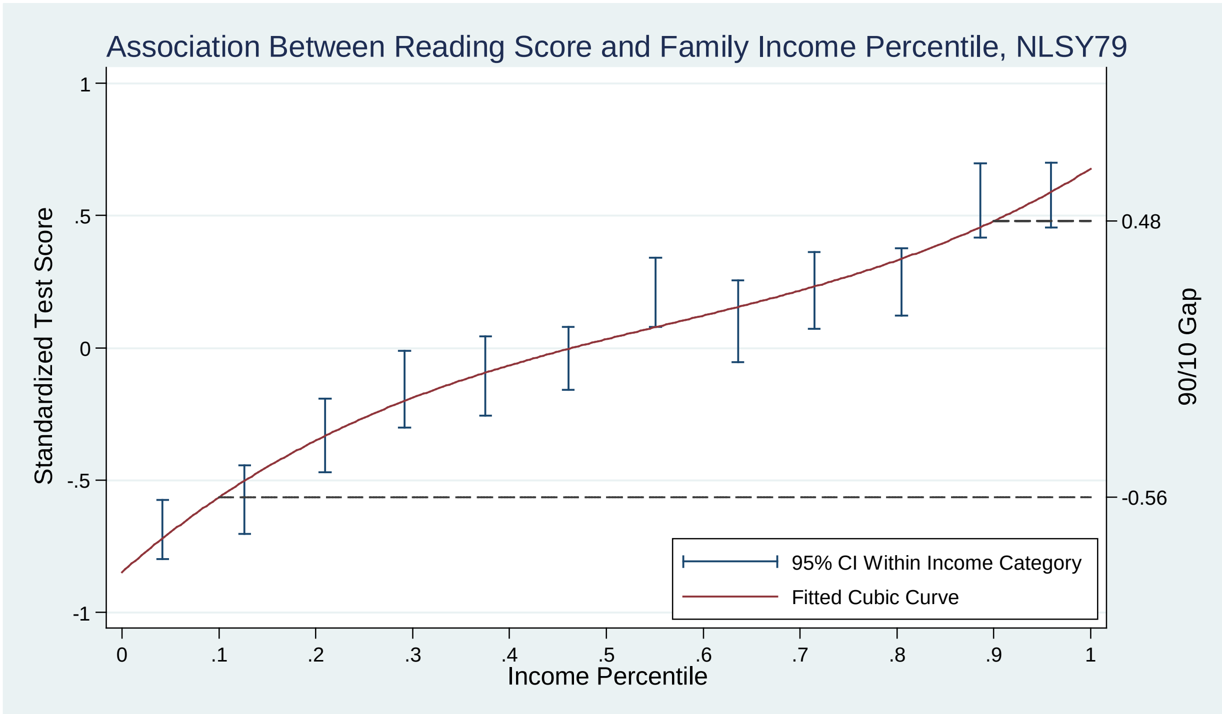

Association Between Reading Score and Family Income Percentile, NLSY79
Standardized Test Score
1
.5
0
-.5
-1
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Income Percentile
0.48
-0.56
90/10 Gap
95% CI Within Income Category
Fitted Cubic Curve

Figure 5.A2

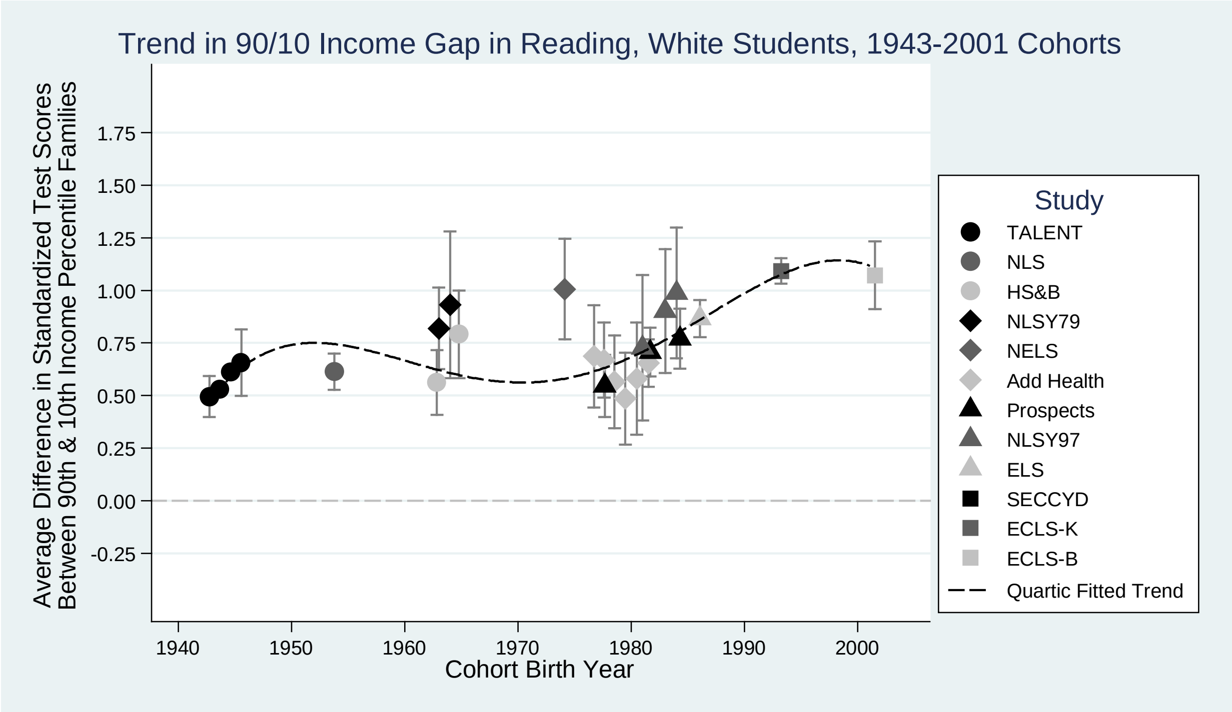


Figure 5.A3

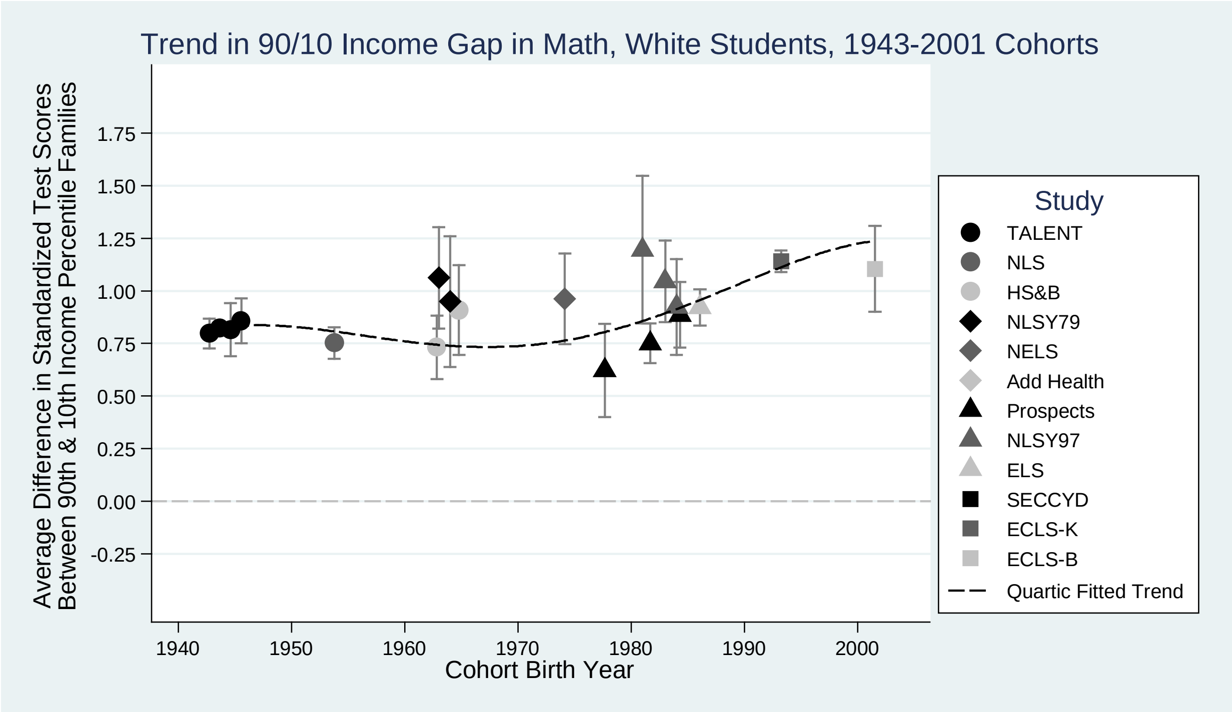


Figure 5.A4

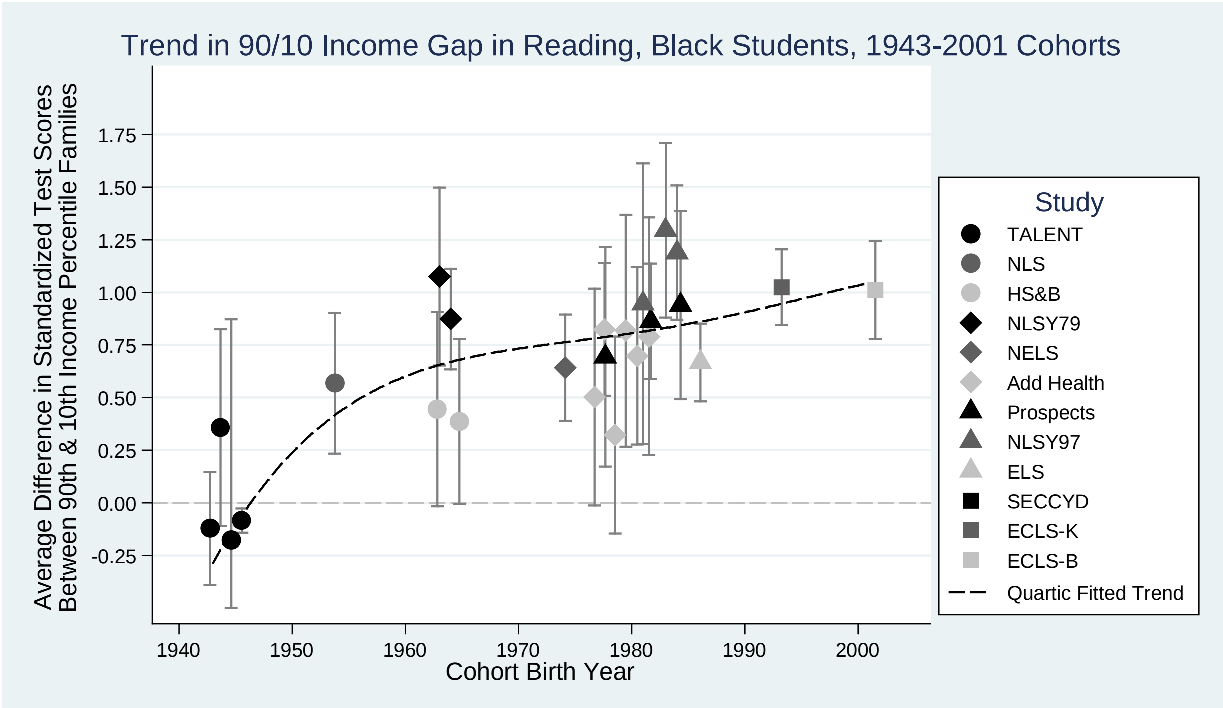


Figure 5.A5

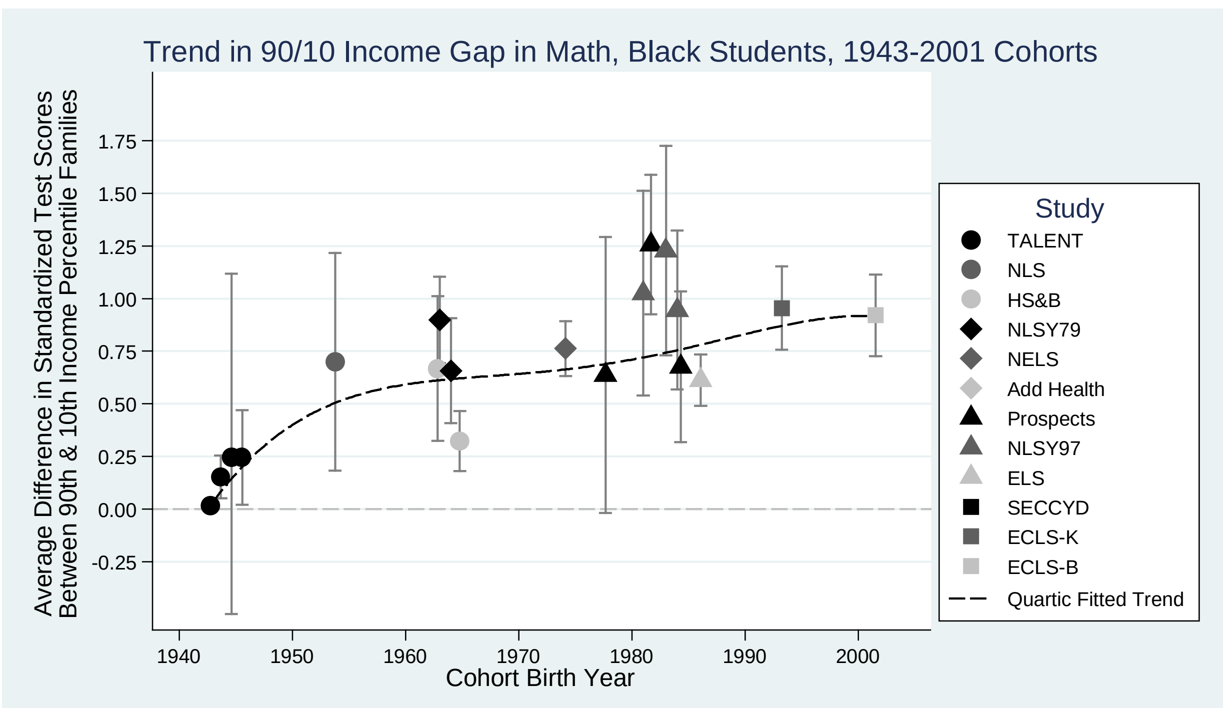


Figure 5.A6

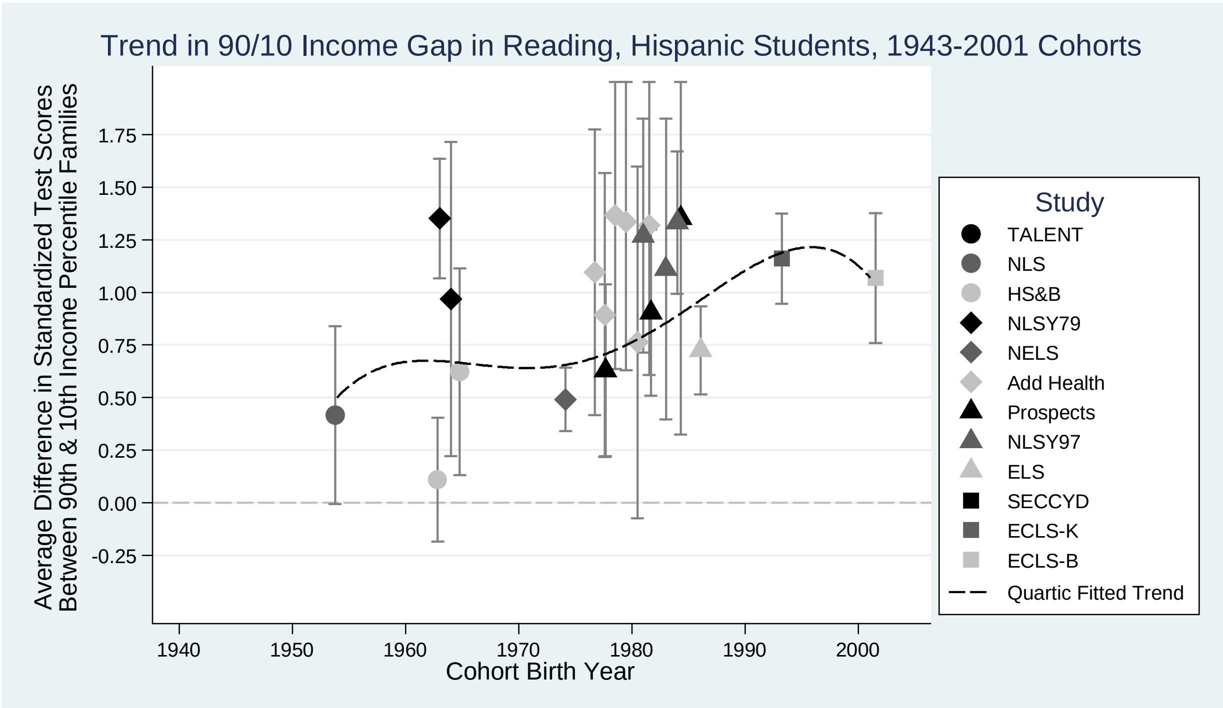


Figure 5.A7

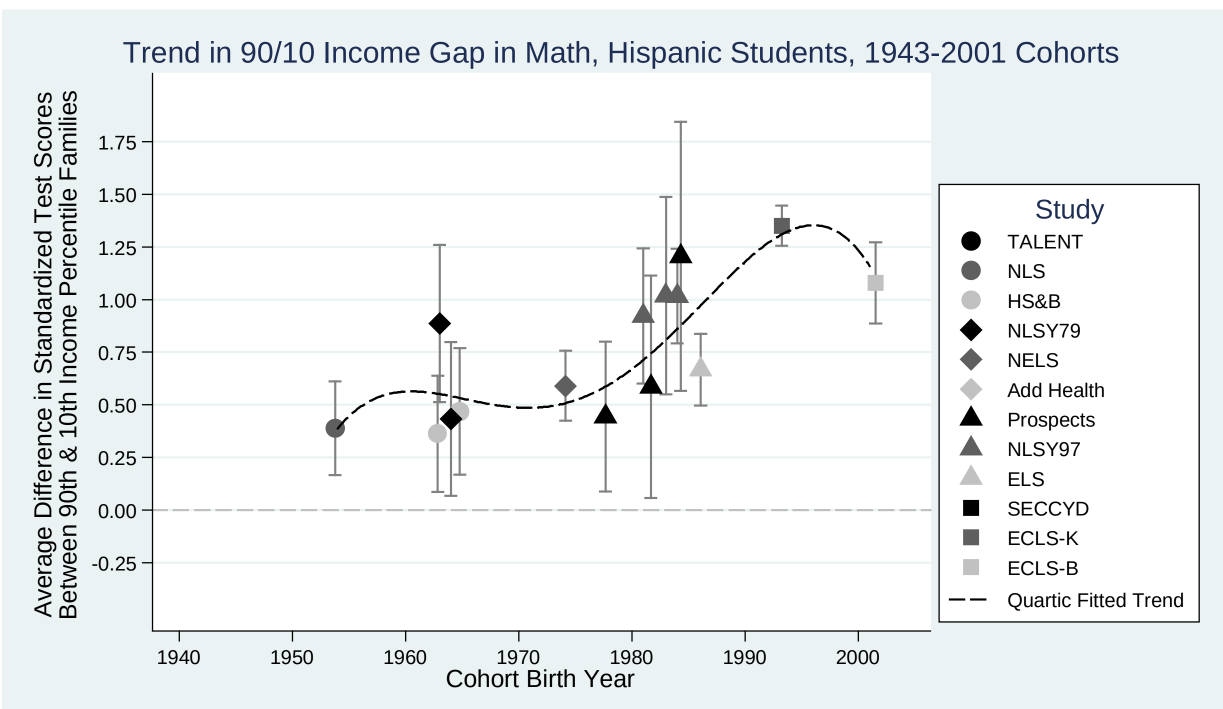


Figure 5.A8

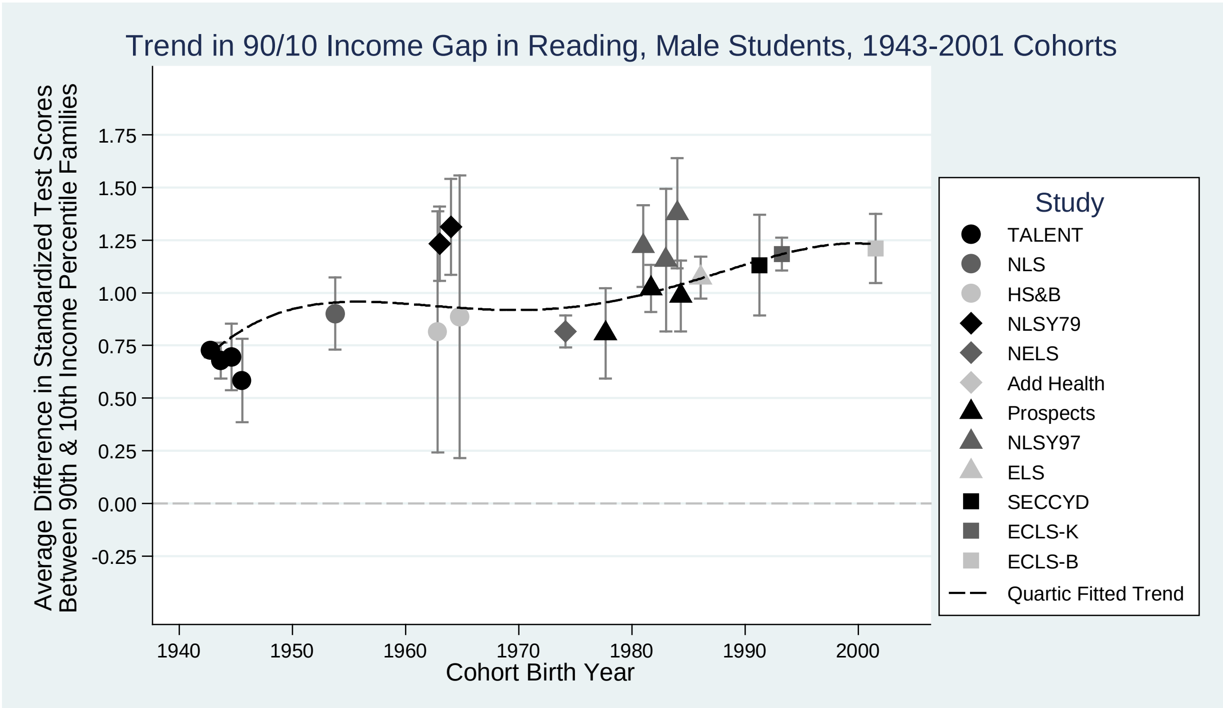


Figure 5.A9

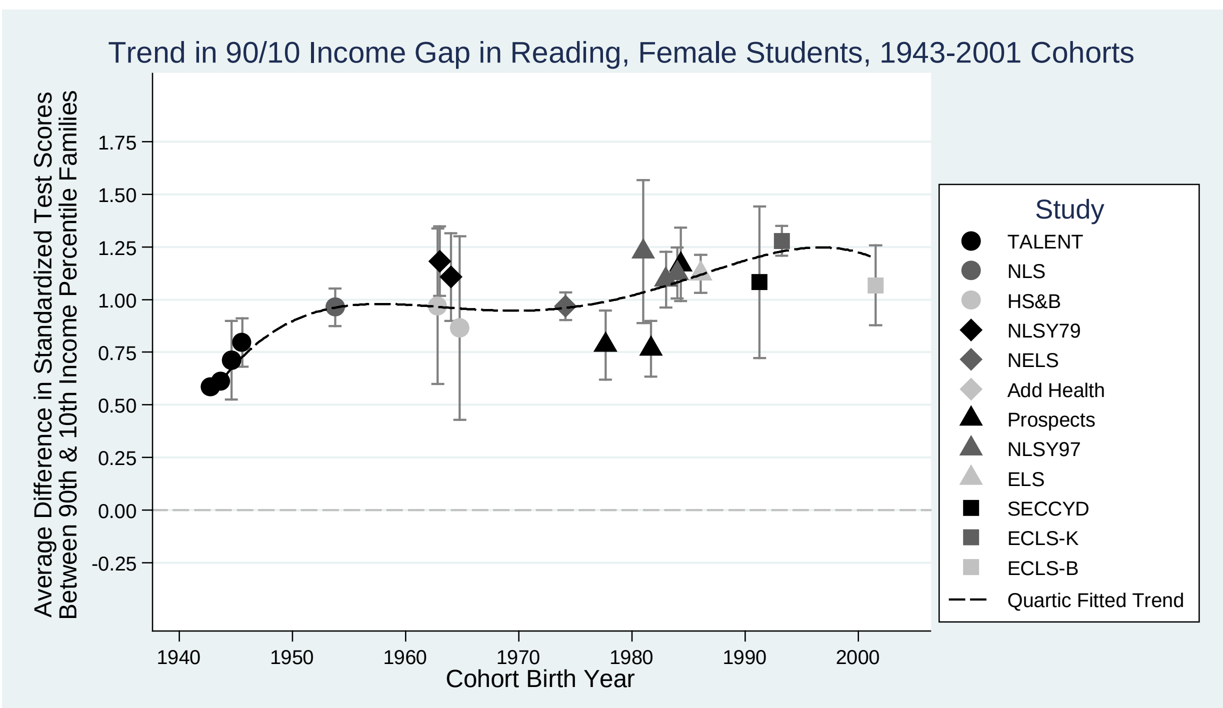


Figure 5.A10

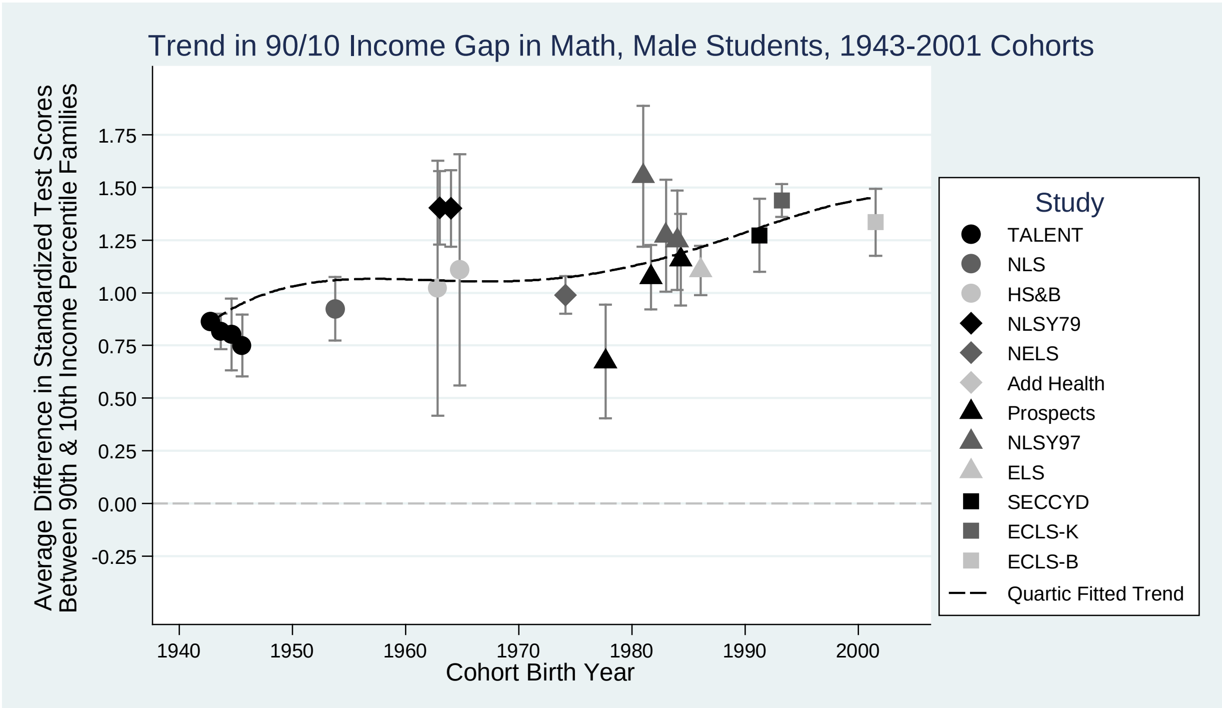


Figure 5.A11

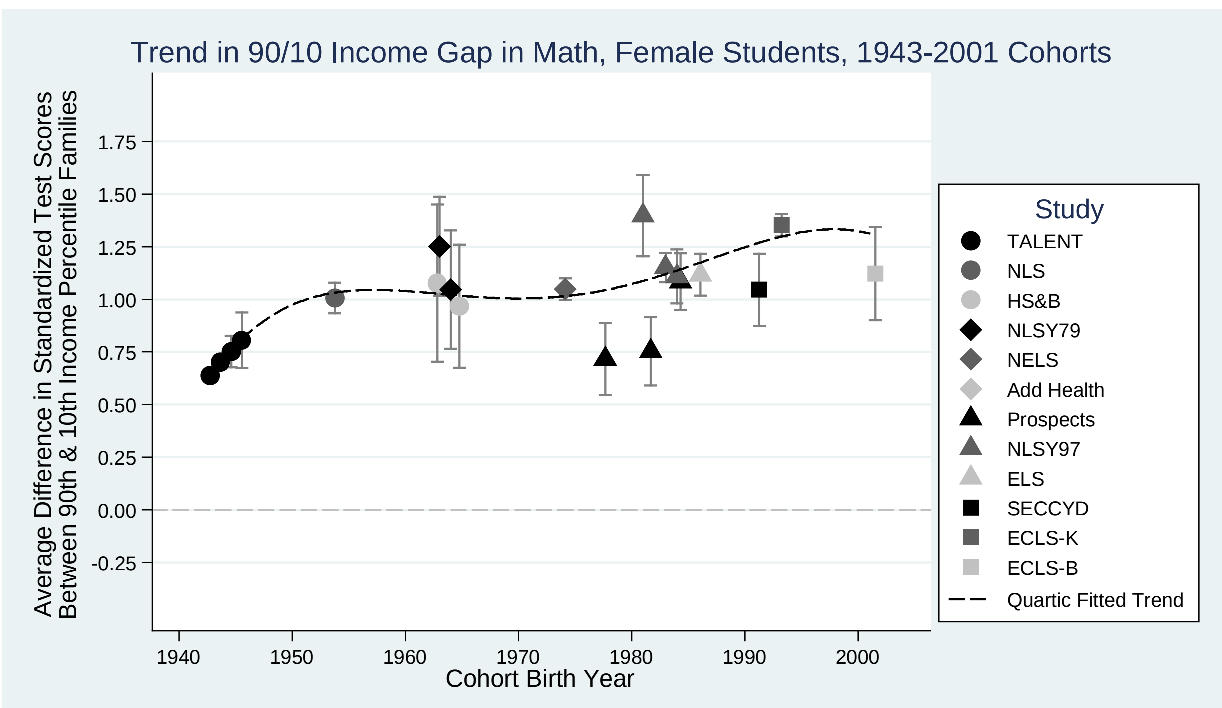


Figure 5.A12

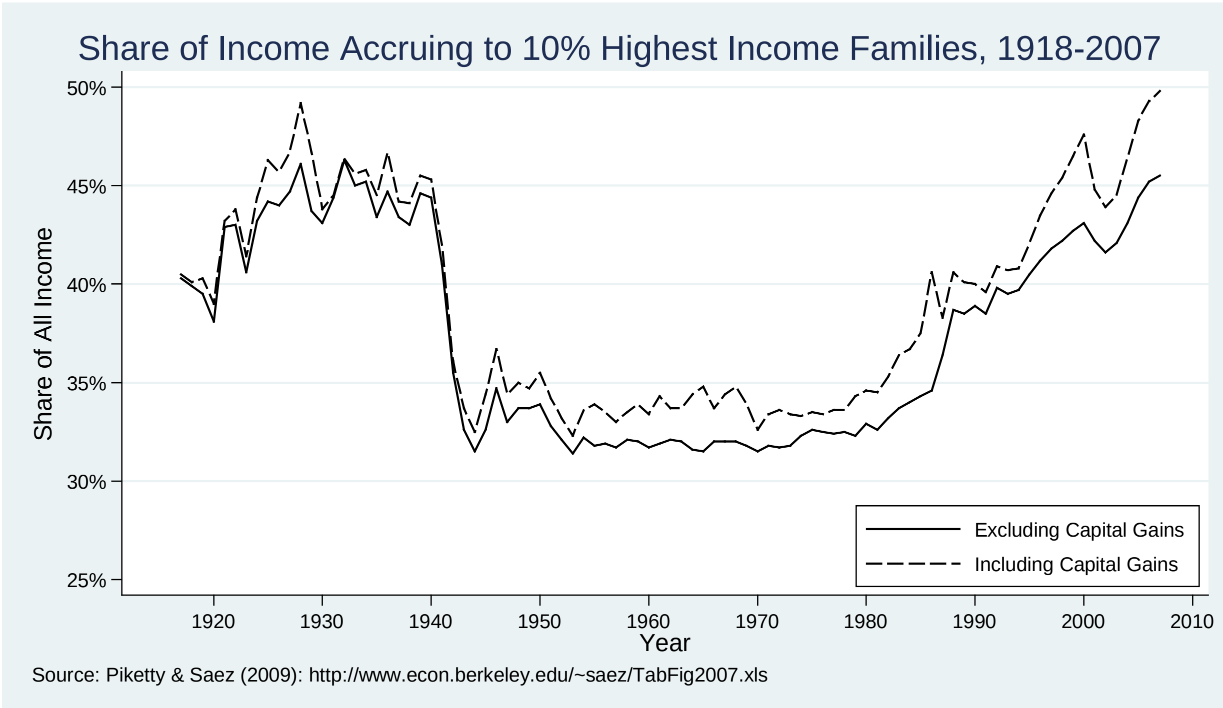

Share of Income Accruing to 10% Highest Income Families, 1918-2007
Share of All Income
50%
45%
40%
35%
30%
25%
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year
Excluding Capital Gains
Including Capital Gains
Source: Piketty & Saez (2009): http://www.econ.berkeley.edu/~saez/TabFig2007.xls

Figure 5.A13

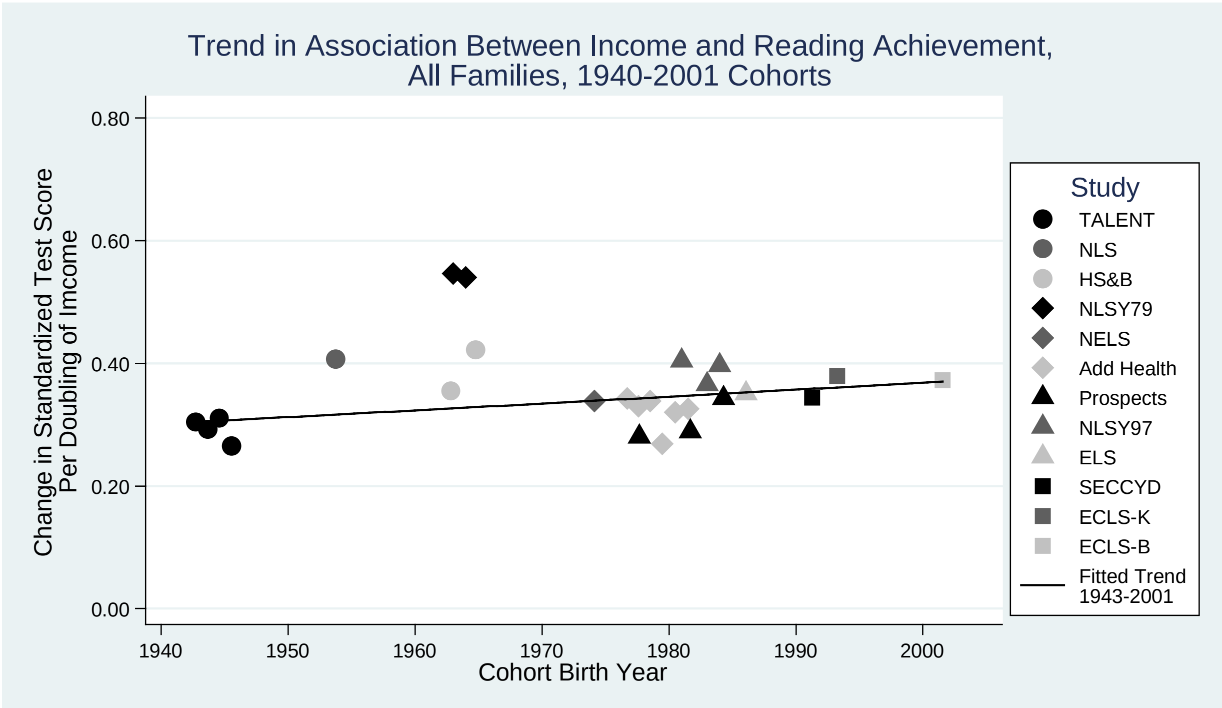


Figure 5.A14

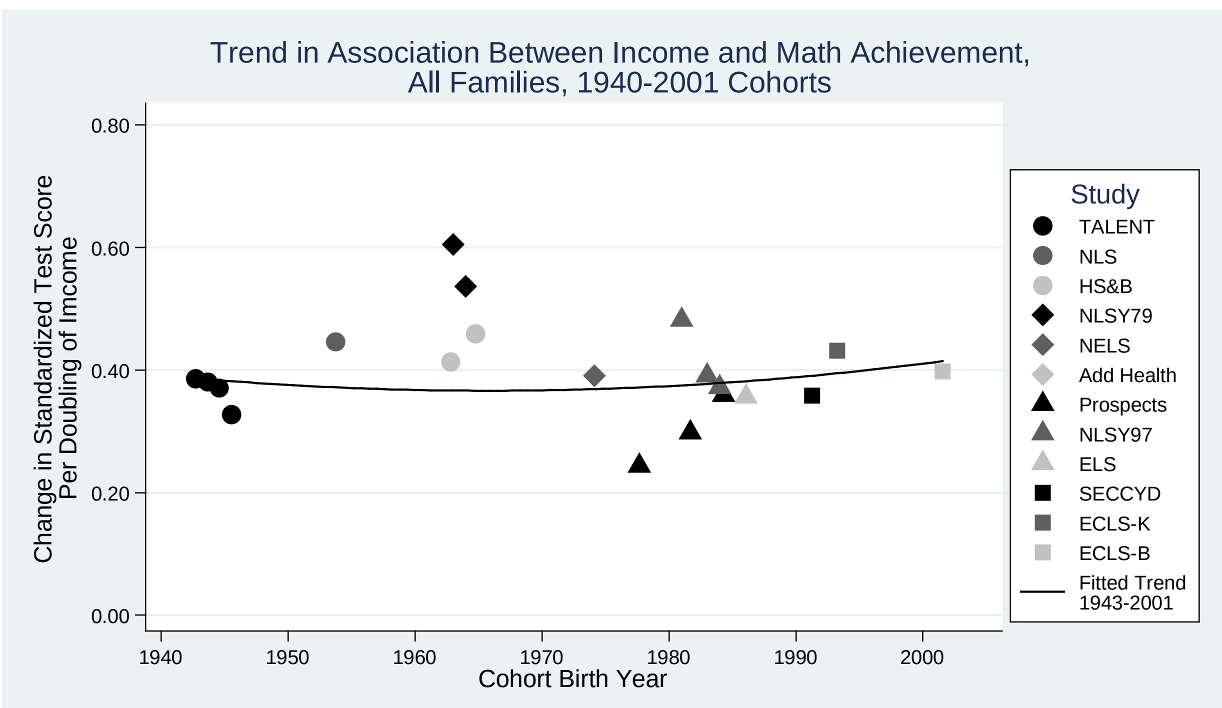


Figure 5.A15

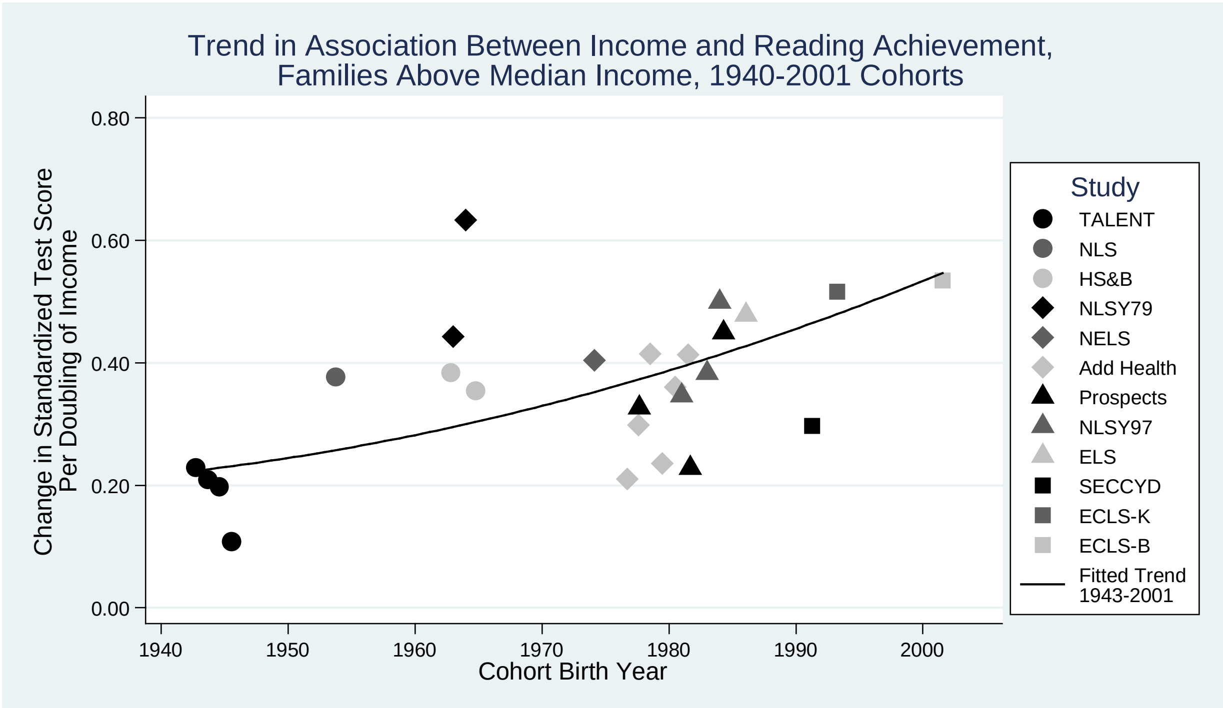


Figure 5.A16

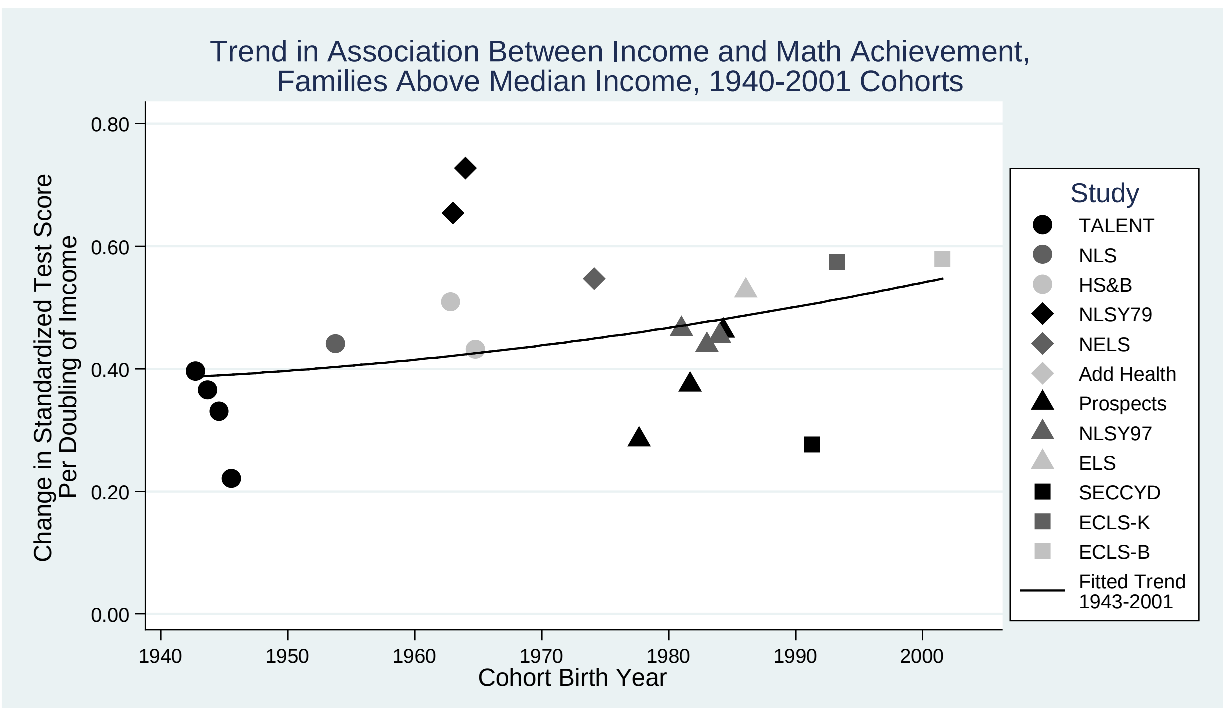


Figure 5.A17

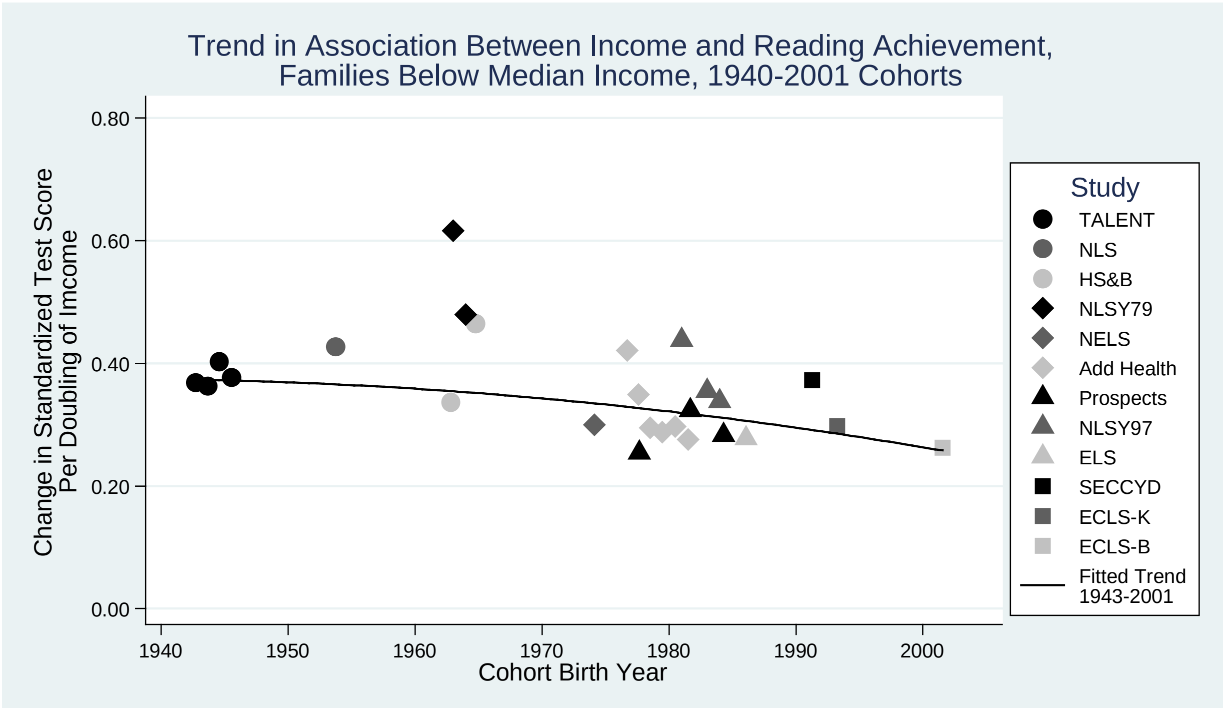


Figure 5.A18

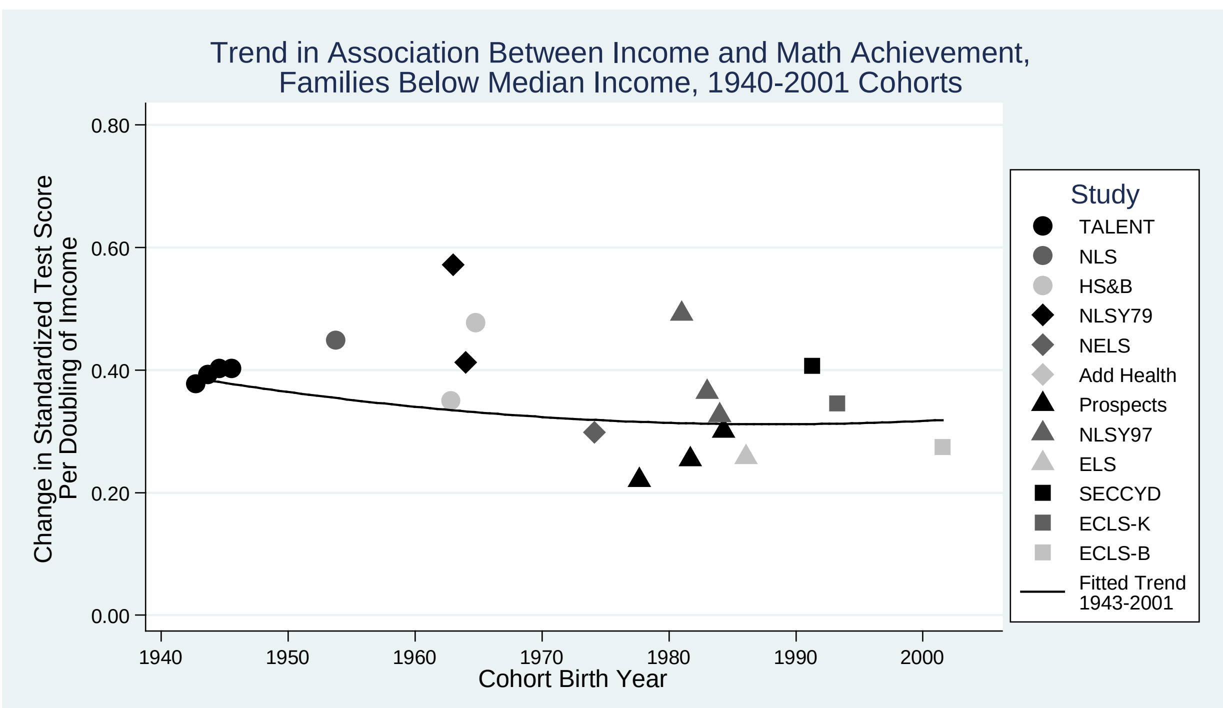


Figure 5.A19: Parental Education Gap in Reading, 1940-2001 Cohorts

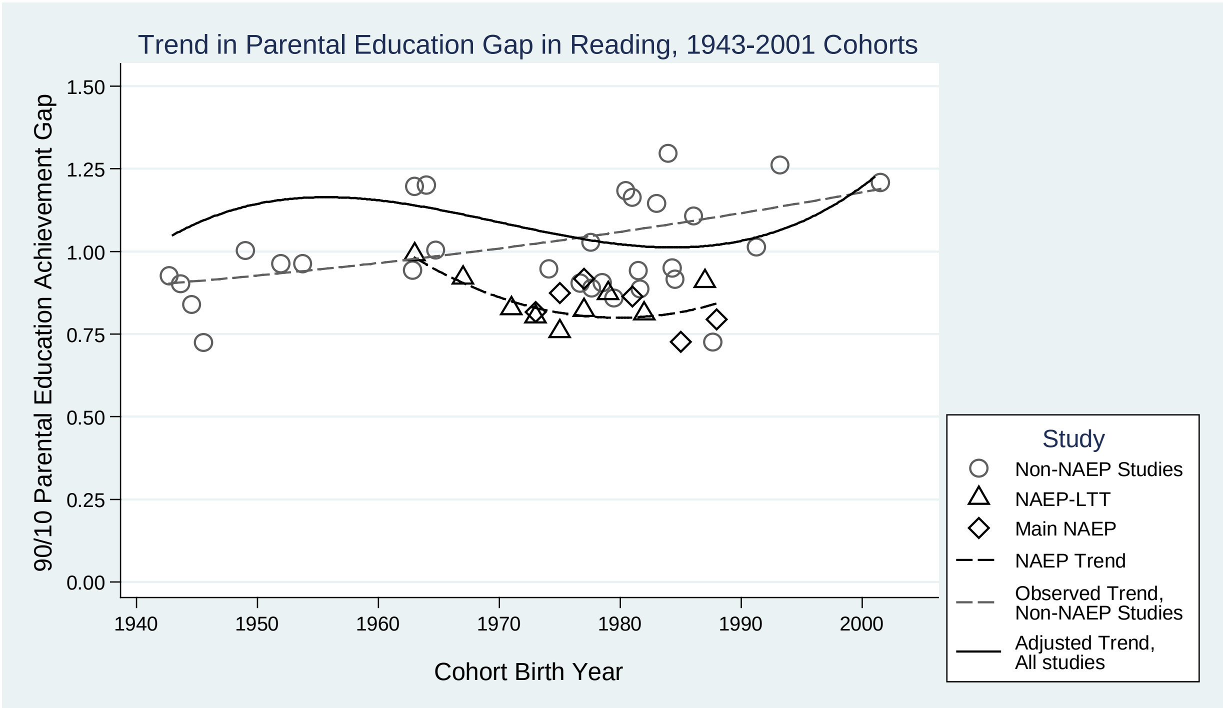
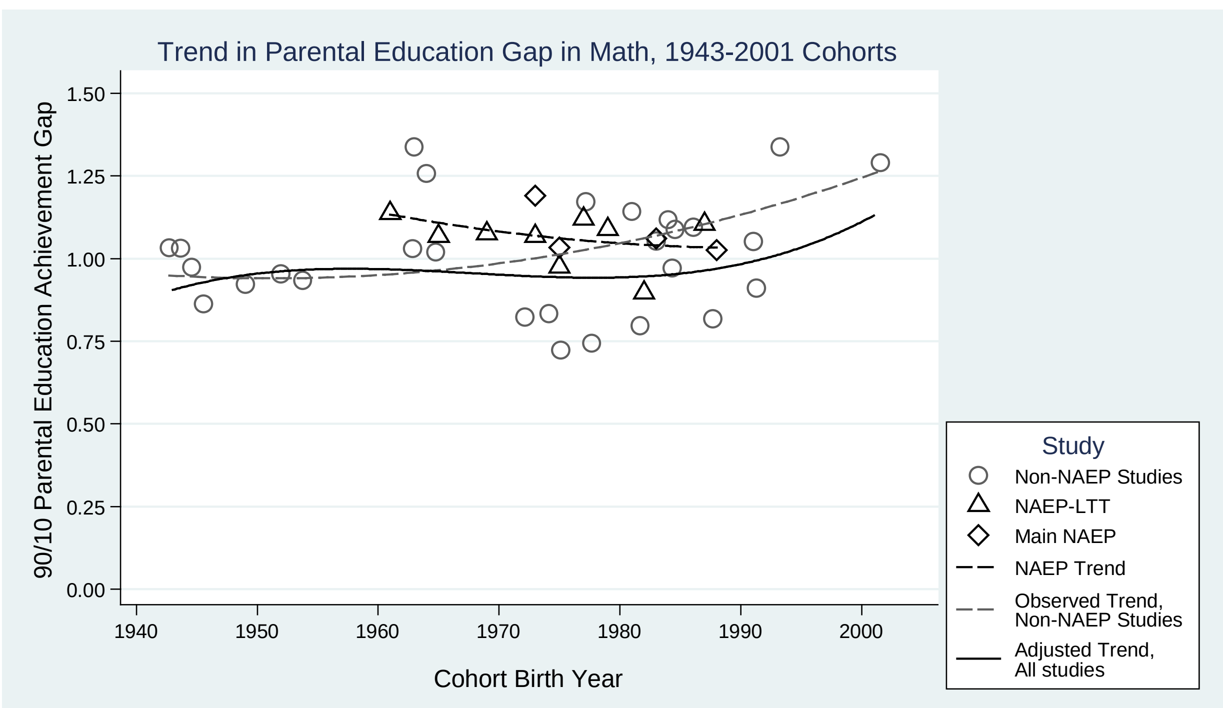


Figure 5.A20: Parental Education Gap in Math, 1940-2001 Cohorts



Note: Figures 5.A19 and 5.A20 display estimated 90/10 parental education gaps from each sample. See Appendix Section 5.A2 for details on the computation of these gaps.