First Draft: March 2019

# "O YOUTH AND BEAUTY:"

## CHILDREN'S LOOKS AND CHILDREN'S COGNITIVE DEVELOPMENT

Rachel Gordon, Robert Crosnoe and Daniel Hamermesh*

*Department of Sociology, University of Illinois—Chicago; Department of Sociology, University of Texas at Austin; Department of Economics, Barnard College, and IZA. The main title is that of a short story by John Cheever. This project was principally funded by the National Institutes of Child Health and Human Development under Grant R01HD081022. We are grateful for core research assistance from Lilla Pivnick and Sarah Moberg as well as additional support and ratings from numerous graduate and undergraduate students. We also thank Simon Burgess, Hannah Hamermesh, Richard Murphy and Jonah Rockoff for helpful comments.

**ABSTRACT**

We use data from the 11 waves of observations of children in the U.S. Study of Early Child Care and Youth Development (SECCYD), 1991-2005, following them from ages 6 months through 15 years. Videos of the children were rated by observers to obtain measures of overall looks at each age. We first examine how these perceptions of beauty change as the children age, finding substantial agreement among raters of a child at each age, with greater agreement for older children. The central part of the study is an examination of how, given their backgrounds (family income, parents' education, race/ethnicity and gender), being better-looking at each age affects changes in the scores on measurements of various objective learning outcomes—mostly achievement tests in math, reading, etc.—the value-added due to looks at each age. First-order autoregressions show that the gains in good-looking children's scores from one age to the next higher age are greater than those of worse-looking children, implying a long-run impact on achievement of at least 0.04 standard deviations per standard deviation of differences in looks. Similar estimates on reading and arithmetic test scores at ages 7, 11 and 16 in the U.K. National Child Development Survey 1958 cohort show somewhat larger effects. These extra gains persist when controlling for teacher ratings of their closeness to study children and maternal ratings of children's behavior. We use results from both data sets to measure the additional economic returns to beauty resulting indirectly from its effects on test scores and hence educational attainment. They suggest that these effects account for 10 to 20 percent of the returns to education.

We find a delight in the beauty and happiness of children. [Emerson, 1871]

## I. Introduction

An already immense and still burgeoning literature has studied the productivity of different inputs into educational production functions, evaluating their effectiveness by examining their valued-added, typically measured in standard-deviation units of changes in scores on various achievement tests. The economic literature goes back at least to Hanushek (1971), with Chetty *et al.* (2014a, b) being just a few of the numerous more recent examples, with an excellent summary of results in Hanushek and Rivkin (2010). Whether experimental (e.g., Fryer, 2011; Abeberese *et al*., 2014) or observational based on administrative data (Aaronson *et al*., 2007; Chetty *et al*., 2014a, b), the general conclusion seems to be that program effectiveness and the differences made by exceptional teachers are small, rarely more than 0.2 standard deviations and often essentially zero. The effects created by the ways in which schools are organized may be even smaller (Dynarski *et al*., 2018).

A much smaller but growing literature has examined the impact of personal beauty on economic outcomes, including earnings (Hamermesh and Biddle, 1994; Harper, 2000; Gordon *et al.*, 2013, and many others), electoral outcomes (King and Leigh, 2009; Berggren *et al*., 2010) and even happiness (Abrevaya and Hamermesh, 2013). The general view is of beauty as a productive characteristic that adds value to a person's performance in a variety of areas (Langlois *et al*., 2000; Hamermesh, 2011). Its effects are not huge, on earnings being somewhere between the equivalent of one-third and a full year of additional education. Given the variances of the distributions of earnings in Western countries, in standard-deviation terms these impacts are, however, at least as large as those found for the long-term effects of the interventions examined in the education literature.

A third literature has examined teachers' expectations and student performance (see Hatfield and Sprecher, 1986, Ch. 5, and Jackson *et al*., 1995, for surveys), although most of the work focuses on how looks affect teachers' perceptions of student ability rather than directly on achievement. A few studies, however, have examined how children's looks are related to their academic performance (e.g., Salvia *et al*., 1977; Talamas *et al*., 2016), but these are quite limited, in that: 1) They use small samples and have few if any controls; and 2) More important, they relate cross-section differences in students' achievements on particular tests to ratings of their looks, thus putting them outside the value-added framework of the education literature in economics.[1]

This study examines the relationship between looks and the value-added, using two very different data sets. Our main focus is on the longitudinal data collected through the American Study of Early Child Care and Youth Development (SECCYD), a panel of over 1000 children who were assessed at 11 different times between ages 6 months and 15 years (between 1991 and 2005). As an attempt to examine the value-added effect of looks on student achievement in a quite different environment and with a different type of data, we also use the 1958 cohort of the British National Child Development Survey (NCDS), which assessed children at ages 7, 11 and 16, and has followed them at various intervals through adulthood.

We cannot identify whether the value-added, as proxied by changes in achievement test results, that we measure is attributable to the child's teacher, his/her parent(s), his/her peers, in-class or outside, or his/her interactions with any of these agents. All that we examine is how the value-added is mediated by the child's looks over the time when the value is being added. Despite this inability, however, we use some proxies describing interactions between the student and the teacher (parents) to examine the robustness of any beauty effects that we observe.

In Section II we first discuss how we measure the beauty of the children in the SECCYD, then move on to analyze patterns of their beauty and how these varied over time. Section III discusses the variables used in the autoregressions of achievement, focusing particularly on the changing variety of achievement measures included in the survey as the children aged. In the next Section we estimate the autoregressions describing value-added by looks in the SECCYD. Section V considers the impact of looks on the value-added in school achievement in the NCDS, while Section VI investigates the possible mechanisms through which looks raise educational value-added. The next Section estimates the extent to

[1]While looking at cross-section effects, and thus outside the value-added literature, Gordon *et al*. (2013) related looks to GPA and other outcomes in the National Longitudinal Survey of Adolescent Health.

which the impact of education on earnings—one of the most widely-examined economic relationships—arises from the impact of looks on educational attainment, first indirectly using the results from the SECCYD and extraneous estimates of the impact of achievement on educational attainment, then directly using the results from the NCDS and additional estimates on those data.

**II. Beauty in the SECCYD**

*A. Assessing Beauty Through Videos*

The SECCYD is a longitudinal study of 1,364 children and their mothers (NICHD Early Child Care Research Network, 2005). It was begun in 1991, when newborns were sampled from hospitals at 10 sites in 9 states. After screening, 89 percent of scheduled one-month interviews were completed. In-person data collections—the "major assessments," which included videotaped interactions, occurred at eleven points: at 6, 15, 24, 36, and 54 months, in grades 1, 3, 4, 5, and 6, and at 15 years. There were videos of from 63 to 93 percent of the initial sample at each assessment (see Table 1). A near majority had videos at all eleven waves (N = 558), and a majority did so at least at ten (N = 782).

Undergraduate research assistants created short slices of video (approximately 7-10 seconds in duration) at each wave of the survey, focusing on the child's face and body. The background setting and other people were blacked out of the videos and the audio was muted, to focus ratings on what the child looked like. This approach is fairly similar to that followed by Benjamin and Shapiro (2009) and can be viewed as a subset of the many studies of the impacts of beauty based on photographs, as opposed to those based on in-person assessments of the subjects' looks, either by interviewers (as in Hamermesh and Biddle, 1994, and Gordon *et al*., 2013) or by people interacting with the subject more frequently (e.g., Harper, 2000).

Undergraduates from the same general birth cohort as members of the SECCYD sample (aged in their early 20s in 2016-18) at two large public universities rated the video clips. Among other things each student was asked to assign ratings from 5 (very attractive/very cute), to 4 (attractive/cute), 3 (about average), 2 (unattractive/not cute) or 1 (very unattractive/not at all cute) in response to the question: How cute/attractive is child/adolescent overall? Each rater had five seconds to rate the subject's overall appearance.[2] In each wave the looks of each subject were assessed by at least ten raters.

The distributions of the raw ratings of overall appearance are presented in Table 2 for the entire sample over all eleven waves. Where a rater looked at fewer than 50 videos in a wave of the SECCYD, that person's ratings were deleted.[3] As is standard in studies of adult beauty (Hamermesh, 2011, Chapter 2), many more people are rated attractive or very attractive than are rated unattractive or very unattractive. Because raters differ in the generosity of their views of the children's/adolescents' looks, each rater's scores were unit-normalized using the rater's own mean and standard deviation within each wave. These rater/wave normalized measures of beauty were used in all subsequent analyses. Appendix Table A1 describes the rating procedures in more detail.

*B. Changing Patterns of Beauty in the SECCYD*

For each subject in each wave (12,045 data points in all) we calculated the mean and standard deviation of their rater/wave normalized individual ratings, creating two variables: 1) The youth/wave mean of normalized ratings and 2) The youth/wave standard deviation of the normalized ratings. For brevity, we refer to these as mean looks and SD looks. Mean looks averaged 0.0015 across all subjects/waves, with a standard deviation of 0.53. SD looks averaged 0.84 across all subjects/waves, suggesting far from perfect agreement among raters. Nonetheless, combining the moderate average intercorrelations with our relatively large number of ratings produced high internal consistency (with Cronbach's $\alpha$ ranging from 0.66 with ten raters of the 15-month-olds to 0.91 with eleven raters of the 15-year-olds, and an average $\alpha$=0.88). Because we used many more raters of each subject's looks than in most other studies (the Wisconsin Longitudinal Study, Scholz and Sicinski, 2015, being an exception), the measured agreement here is higher than in most other studies.

[2]For much more detail on how the videos were created and how the coders were instructed, see (Gordon *et al*., 2018).

[3]These were fewer than 1 percent of all ratings. Including them hardly changes the distribution in Table 2.

Table 3 shows the averages (across all individuals) of each child's mean looks by gender and by wave of the survey in the first two columns, and the averages of each child's SD looks at each wave and gender in the second two columns.[4] The table demonstrates several regularities in how the children's/adolescents' looks are viewed by the raters. Girls' looks are consistently rated higher on average than boys'. This is different from the results of most research on adults, where there is little difference on average by gender. The differences here are quite small, however, with the average girl being in the 55th percentile of looks in the overall sample, the average boy being in the 44th percentile.

The gender difference in the average ratings of looks generally rises over the first 15 years of life, although not monotonically, to the point where at age 15 the average teenage boy's looks place him at the 39th percentile in the sample, while the average girl's place her in the 61st percentile. These gender differences do not arise because people find it more difficult to rate boys' looks. On the contrary, in seven of the eleven waves the average SD of looks is significantly less for boys than for girls, while only at age 15 is the opposite true.

Our focus is on the impact of looks on child development, as measured by value-added in students' achievement over time; and since we know that the latter is affected by income and demographic differences, it is crucial to examine whether these are also correlated with the youth/wave aggregated normalized ratings of looks. The SECCYD contains information on the race/ethnicity of the child, the income of the child's parents when the child was born, and indicators of the mother's educational attainment and that of the better-educated parent. Appendix Table A2 presents statistics describing these control variables.

The SECCYD sample was randomly drawn from hospital births in each site, and well matches demographically the catchment areas of those hospitals (NICHD SECCYD Steering Committee. 1993). The sample also tracks well the distribution of Americans with children ages 0-2 in 1990, with some differences reflecting the geographically-restricted sample. The racial/ethnic distribution matches perfectly

[4]Note that the standard deviations of the averages of the normalized ratings are less than one, because we are averaging across the positively but not perfectly correlated normalized ratings.

the fraction African-American in the relevant population; but Hispanics are twice as frequent in this sample as in the population of parents with children ages 0-2 in 1990 (12 percent vs. 6 percent), and there are commensurately fewer non-Hispanic Whites. The income brackets used by the SECCYD are similar to income quartiles in the 1990 Census, and reported incomes in the sample of parents here are somewhat higher than those in the population of similarly aged adults (e.g., nearly one-fifth versus one-tenth in the top quartile; about one-quarter versus one-third in the bottom quartile).[5] The distribution of educational attainment indicates, consistent with the distribution of income, that parents of the children in this sample are more educated than the average parent with a newborn/toddler in 1990 (two-thirds of the sample versus only 43 percent of the population having more than a high school diploma).

Table 4 presents the results of estimating how demographic characteristics affect the mean and SD of looks for the entire sample and for girls and boys separately. Overall, the associations are quite small, with only 1 to 3 percent of the variance explained. Relative to Hispanics, children in all three other racial/ethnic groups receive average ratings that are lower, although only with meaningfully large differences relative to non-Hispanic blacks. There is, however, less agreement among raters about the looks of blacks and other non-Hispanic children, as shown in the final three columns of the table. Raters gave slightly higher ratings to children from higher-income families, although not statistically significantly so; and there was (insignificantly) more agreement among raters about the looks of those children. Overall, and especially because of the differences in beauty ratings by gender and by race/ethnicity (particularly for non-Hispanic blacks), these results underscore the importance of accounting for demographic differences in describing the impacts of looks on educational value-added.

**III. Outcome Measures in the SECCYD**

The SECCYD included various tests of the child's/adolescent's development in each wave. Because many measures were designed and selected to be age-appropriate, none was used for all ages

[5]The distributions of race/ethnicity and educational attainment are from the 1990 CPS-MORG; the income distributions are from the 1990 Census.

between 6 months and 15 years. Thus we cannot simply examine the value-added using the same assessment as the child ages. Rather, we use various measures, concentrating on those which are present in as many waves as possible and which represent objective evaluations. As checks on the validity of our estimates, we experiment by estimating the impacts of looks and other measures on alternative assessments in each wave.

Table 5 lists all the outcome variables that we use in the results presented in the text tables. For each we list the variable name, a description and the waves in which it was used, and its mean, standard deviation and range. The most frequently provided assessment, the Woodcock-Johnson Applied Problems Standard Score (*WJAPSC*) Revised Version (Woodcock and Johnson, 1989), is a math subscale from a battery of tests designed for standardized administration by trained staff to assess achievement from early childhood through old age. It has been used very rarely by economists (with Akresh and Akresh, 2011, and del Boca *et al*., 2017, being rare exceptions), but is standard among educational psychologists (http://achievement-test.com/testing-options/woodcock-johnson-iii-tests).[6] We use the standard score which the SECCYD study staff looked up in tables created by the test developers using a norming sample. As frequent as the *WJAPSC* and overlapping it in availability in three of the five waves for which it is available, is another set of achievement scores, the Academic Skills Rating Scale (*ASLL*). The ASLL is the average of ten items that teachers rate on a scale of 1 (Not Yet) to 5 (Proficient) to reflect children's language and literacy skills (e.g., conveying ideas clearly, understanding stories read aloud, composing multi-paragraph stories). Because it seems less likely to be objective than the WJAPSC, we use it only when the latter is unavailable.

These measures cover student achievement from Wave 5 (age 54 months) through Wave 11 (age 15) but are not administered to toddlers and pre-school students. For them (Waves 1-4) we use age-appropriate standardized measures administered by SECCYD study staff, concentrating on the Bayley

[6]There are several other assessments in the Woodcock-Johnson battery with which we experimented but which we do not report here. We do, however, present the results using one set in an Appendix Table.

Mental Development Index (http://www.healthofchildren.com/B/Bayley-Scales-of-Infant-Development.html ) in Waves 2 and 3 (ages 15 and 24 months respectively), and the Bracken School Readiness Composite (https://www.pearsonclinical.com/childhood/products/100000165/bracken-school-readiness-assessment-third-edition-bsra-3.html) at age 36 months (the earliest for which this measure is age-appropriate).[7] Finally, IMPRSO, used at Wave 1 (age 6 months) is impressionistic rather than based on any formal testing or assessment.

As the final three columns of Table 5 demonstrate, the assessments all have different scoring systems and, although available for all the subjects in the SECCYD who remained in the study at any wave, are not directly comparable. To enable comparisons, we normalize each measure, separately at each wave of the SECCYD. The outcomes that we examine at each wave are thus the normalized scores of the child's/adolescent's achievements on the particular assessment measure.

**IV. Looks and Educational Value-Added During Childhood**

The general models to be estimated are:

(1) $S_{it} = \alpha B_{it-1} + \beta S'_{it-1} + \gamma X_{i0} + \varepsilon_{it}$ , t = 2,…, 11 ,

where S is the normalized score on some educational assessment, S' is either the same assessment mode or one closely related, B is the mean looks of child i at time t-1, X is the set of controls describing family and parental circumstances at the child's birth, t-1 is the time of the previous assessment, and ε is the usual disturbance term. Because the waves are not spaced evenly over the child's 15 years, the lag in (1) can be anywhere from 9 months to 4 years (between Waves 10 and 11). We estimate the model for each cross-section separately (each wave), then pool the data for all the waves.

Table 6 presents the results for Waves 2-6, Table 7 the results for Waves 7-11. We show both the assessment measure used as $S_t$ and that used as $S'_{t-1}$. Where the same measure is available in two consecutive waves, as at 24 months and Grades 1, 3 and 6, we use that measure in the equations underlying these tables. Where a standardized test, rather than a teacher assessment, is available, we use that (all waves

[7] While both sets of assessments are standard in educational psychology, they have rarely been used by economists (but see Duncan *et al.*, 2007; Rubio-Codina *et al*., 2015).

but Wave 1, where IMPRSO is based on observers' impressions, and Wave 10). In each case we only show the estimates of the expanded versions of (1) that include the entire vector of covariates X. Of the ten estimates of α in these Tables, eight are positive, of which four have t-statistics above one. (The estimates of α from equations excluding the vector X are uniformly more positive and more likely to be statistically significantly greater than zero.) Remembering from Table 3 that the standard deviation of mean looks is 0.53 (averaging boys and girls), the estimates of the impact of a one standard-deviation increase in average beauty on value-added in the educational assessment range across the ten waves from -0.02 to 0.09 standard deviations, with an average estimated impact of a 0.03 standard-deviation increase in achievement. This is at least in the middle of the broad literature on the estimated value-added of a good teacher, and above a recent estimate of the impact of disruptive peers on test scores (Carrell *et al.*, 2018).

Replicating and extending prior studies using the SECCYD (e.g., Crosnoe *et al.*, 2010; Vandell *et al.*, 2010), we find that persistence in achievement is quite strong (except at Wave 2, where the lagged variable is an overall rating of the mother-child interaction, rather than child achievement) and generally rises as the children age. Especially before they enter grade school, the value-added for girls exceeds that of boys. What is most remarkable is the importance of race/ethnicity and parental income on the change in scores—on the value-added by education and whatever else increases children's achievements—between these assessments. The value-added among non-Hispanic Blacks is uniformly negative compared to that among otherwise identical Hispanics, which in turn is, with one exception, uniformly less than that among non-Hispanic Whites. While the value-added does not rise monotonically with the family's income at the child's birth, children born to families in roughly the top income quartile generally see greater improvements in their test scores than children born to families of the same race/ethnicity in the lowest income quartile.

Even at Wave 2, before there has been substantial sample attrition, the number of observations used to estimate (1) is not large. To increase power and precision and provide sufficiently large sample sizes for gender-specific models, we pool the estimates for the ten waves, using the same measures of $S_t$ and $S'_{t-1}$ as in Tables 6 and 7 (and cluster the estimated standard errors on the individual child). We show the results of estimating the pooled equations in Table 8, for the entire sample and for girls and boys separately, and without and with including the vector X. Examining the estimates of the impact of better looks on the value-added between assessments for the entire sample, when the vector X is excluded the impact is an additional immediate impact of 0.05 (.101*.53) standard deviations for each one standard-deviation increase in mean looks. Including the entire vector drops the estimated immediate impact to 0.03 standard deviations (not surprisingly, the average of the estimates from the previous two tables). In either case, however, the estimated impacts are statistically significantly nonzero. The long-run impact of a one standard-deviation increase in looks on these scores is 0.11 (.101*.53/[1-.528]) in the estimates without covariates, 0.04 (.045*.53/[1-.420]) in those including covariates.

The bottom rows of Table 8 decompose the sources of the declines in the estimated effects of standardized beauty on gains in achievement using Gelbach's (2016) method. For both sexes pooled, and for boys and girls separately, over half of the decline is due to the addition of the race/ethnicity indicators, with parents' education generating one-fourth of the decline in the entire sample, and household income at the child's birth never accounting for more than one-sixth of the drop. Given the relative differences in standardized beauty by race/ethnicity shown in Table 4, the results of this decomposition are not surprising

We cannot reject the hypothesis that the estimated impacts of looks on value-added are equal between girls and boys. Nonetheless, whether or not the vector X is included, the estimated impacts are greater among boys than for girls, consistent with results in the majority of the literature on gender differences in the effects of looks on labor-market outcomes among adults (summarized in Hamermesh, 2011, Chapter 3). Confidence in the SECCYD gender difference is reinforced because the standard errors reflect nearly identical precision of estimation of the effect for boys and for girls.[8]

In the pooled estimates with the vector X included, the basic conclusions from the previous tables remain, but the results generally become statistically significant, and the variations across waves are

[8]Estimating the pooled model separately for white non-Hispanics yields slightly smaller estimated effects of beauty on value-added. The estimates for the smaller samples of other children produces slightly larger estimates than those shown in Table 8.

smoothed out. Thus the value-added rises consistently, other things equal, as we move up the distribution of parental incomes at birth. Similarly, and consistently, the value-added among African-American children is significantly less than that among Hispanics, which is less than that among non-Hispanic Whites, whose value-added is not statistically different from that of the small number of non-Hispanic members of Other races included in the SECCYD. Similarly, the average value-added for girls at each wave slightly exceeds that for boys of the same race/ethnicity and family income background.

One might be concerned about the robustness of the beauty effects to specification errors resulting from unobservable and thus excluded variables. Following the method of Oster (2019), we can calculate to what extent selection on unobservables would need to be correlated with that on observables if inclusion of the former were to increase the $R^2$ by 30 percent. For the fully-specified equations in Table 8, selection on unobservables would need to be greater than that on the observables to vitiate the robustness of the estimated impacts of beauty on the value-added.

Given the different assessments at each wave of the SECCYD, numerous additional regressions could serve as robustness checks on the results in Tables 6-8. For example, pooling all the assessments that include the variable WJAPSC and re-estimating the full version of (1), yields an estimated impact identical to the 0.045 shown in Column 4 of Table 8 (although with less than half as many observations, the estimate is barely significantly positive). While some of the robustness experiments yielded very tiny estimated effects of mean looks on value-added, the majority produced results that were quantitatively like those in Table 8. Appendix Table A.3 reports on some of these.

## V. A Re-Assessment Using British Data

While the results in the previous sections provide remarkable evidence of the role of looks in affecting student achievement, they are clearly specific to the timing of the SECCYD, its location (selected sites around the U.S.), the peculiarities of the samples selected, and the measurement of the children's beauty by assessments of videos of them at various ages. This is an acceptable way of assessing looks; and our using multiple raters adds to its reliability. But it is only one such way, with direct assessments by an interviewer or observer of a live subject, or single or multiple assessments of photographs, used more frequently in the literature. To examine the basic idea—whether and to what extent students' appearance affects their intellectual progress, conditional on other measures including family background—using a different method of assessing looks, we examine outcomes in the 1958 cohort of the British National Child Development Study (NCDS).

The NCDS is one of several longitudinal data sets that followed every child born in the United Kingdom during a single week, in the case of the 1958 cohort, during the first week in March (http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=724&sitesectiontitle=National+Child+Development+Study). In this study the child's teacher at age 7 rated his/her looks, in response to the question: "Which best describes the student?", with answers attractive, unattractive, abnormal feature, looks underfed or scruffy and dirty, with an excluded category of none of the above.[9] We discarded the tiny minority of students (2.5 percent) who were viewed as underfed or scruffy and dirty, and classified those viewed as attractive as good-looking, those viewed as unattractive or with abnormal feature as bad-looking, and all others as average-looking at age 7. Similar ratings were provided by the child's teacher at age 11.[10]

The means of these indicators of appearance are presented in Rows (1) and (4) of Table 9. The majority of students are viewed as good-looking, with only a small fraction, less than ten percent, classified as bad-looking. Compared to the multiple ratings of videos by unknown others in the SECCYD data, these single ratings by teachers who knew the children are weighted even more disproportionately toward viewing the children as good-looking.

The NCDS also records the results of students' achievements on objective reading and math tests at ages 7, 11 and 16. The arithmetic test is the standard Southgate test, while the reading comprehension test at age 7 and the reading comprehension and math achievement tests administered at ages 11 and 16 are purpose-constructed for the NCDS. The tests at ages 11 and 16 are very similar in construction. The means

[9]The looks assessments in these data were used by Harper (2000) to examine the impacts of looks on earnings, and by Abrevaya and Hamermesh (2013) to study the effects on earnings and on happiness.
[10]The data set does not indicate whether the child had the same teacher at ages 7 and 11. Practices both of teachers specializing in a grade, and teachers moving up with the student through primary school, existed in the U.K. in the 1960s.

and standard deviations of the raw scores in this sample on each of the six tests are presented in Rows (2) and (3), and (5)-(8), of Table 9. The heterogeneity in scores is substantial in each case, with coefficients of variation much larger on mathematics than on reading tests.

As before, we estimate what are essentially autoregressions describing a test score at some at t (t = age 11 or 16) as a function of the teachers' assessments of the children's looks and his/her test score at age t-1 (where t-1 represents scores and beauty ratings at age 7 or 11). Because the tests differ between ages 7 and 11, because the distributions of scores differ across the three periods in which the students are observed, and in order to compare the estimates here to those in the previous sections, we unit-normalize the scores on each of these tests in all the regressions. In each autoregression we also include an indicator of the child's gender and a large vector of indicators of the social class of his/her father at time t-1. As is common in U.K. data, no information is available on race/ethnicity.

The estimates of the autoregressions describing both reading and math scores at ages 11 and 16 are presented in Table 10. The differences in the changes in test scores between the good- and bad-looking children are uniformly statistically significantly nonzero, 0.222 and 0.163 standard deviations for reading, 0.326 and 0.084 standard deviations for math. These essentially measure the value-added to the student's learning by his/her appearance during the four (five) years since the previous test, independent of other factors that affect the value-added. Even the smallest of these is larger than many of the estimates in the literature of the impact of large increases in teacher quality on value-added (e.g., Hanushek and Rivkin, 2010, and Aaronson *et al*., 2007). The largest of them, on math in the early years, is larger than any estimate of the impact of high-quality teachers on students' learning. Given that the value-added by looks is greater in math than in reading between ages 7 and 11 (although less between ages 11 and 16), the greater persistence of math than of reading achievement suggests a larger long-term effect of looks on math than on reading scores.[11]

[11] The effects are entirely due to the children's looks, not their body types. While a higher bmi at age 7 (11) is associated with a significant, albeit slight extra value-added in test scores, the correlations between bmi and the looks variables never exceed 0.10. Adding bmi to the specifications thus has essentially no impact on the estimated coefficients on

As the table shows, except for math scores at age 11, given their looks, the value-added is significantly less for girls than for boys. Separate autoregressions (unreported) by gender yield similar estimates of the effects of looks and of the lagged terms, showing that the negative effects in the Table for girls do not stem from correlated gender differences in the impacts of other factors. As in the estimates using the SECCYD, the F-statistics for the vector of indicators of father's social class show that this measure has important effects on value-added, with greater increases if the child's father was in a higher social class.[12]

Given that the method of assessing looks is completely different from what we designed for the SECCYD, the impacts of these ratings on value-added in school cannot be directly compared to those of the measures used in the previous sections. In relating them to outcomes and using the averages of the distributions of looks at ages 7 and 11, a move from the excluded category to being viewed as good-looking is equivalent to a move from the 25th percentile of looks (the mid-range of average-looking students) to the 70th percentile (the mid-range of good-looking students). In terms of a unit-normal variate, this is equivalent to an increase of 1.20 standard deviations of looks. A move from being viewed as bad-looking to being viewed as good-looking is equivalent to a move from the 5th percentile of looks (the mid-range of bad-looking students) to the 70th percentile (good-looking students), an increase of 2.19 standard deviations of looks.[13]

Applying these equivalences to the estimates in Column (1) of Table 10 yields the result that moving from average- to good-looking generates a 0.07 standard-deviation increase in reading test scores between ages 7 and 11 per standard-deviation increase in looks. Moving from bad- to good-looking yields

---

the beauty terms. The correlations of looks and bmi among adults are also very low (Oreffice and Quintana-Domeque (2016).

[12]Having a father in a higher social class is generally related to greater value-added in test scores, as in the SECCYD. Because aggregating the eight (seven in the age 16 regressions) into three or even four classes discards information and consistently yields a lower R-bar squared, we do not report results based on aggregated social classes.

[13]With region of the UK being related to looks and mobility (Abrevaya and Hamermesh, 2013), we added a vector of indicators of region at age 11 to the specifications. While the vector itself was highly significant (with Scotland, the Southeast and the South having the largest positive effects on value-added), changes in the estimated effects of looks were minuscule in all cases.

an increase in reading test scores over these four years of 0.10 per standard-deviation increase in looks. Based on the estimates in Column (2) both improvements in looks increase reading test scores between ages 11 and 16 by 0.07 standard deviations per standard-deviation improvement in looks. The analogous figures for math test scores between ages 7 and 11 are 0.10 and 0.15 standard deviations per standard-deviation rise in looks; between ages 11 and 16 they are both 0.04 per standard-deviation increase in looks.

The measurement of looks in this data set is totally different from that we created using the SECCYD. Nonetheless, calculating the estimated impact of a one standard-deviation change in looks on the value-added between assessments yields effects that are larger, but not hugely different from those produced using the other data set.[14]

**VI. Channels of Causation of the Beauty Effect on Educational Value-added**

In the SECCYD, the interpretation of the role of beauty on value-added as possibly being causal is strengthened because looks in the base (lagged) period are measured by outside observers, not by anyone who might have a role in determining the increase in test score from one period to the next. In the NCDS a causal interpretation is arguably also strengthened because looks are an assessment by the teacher in some early grade, and since the increase in test scores occurs over the four- or five-year time period during most of which the student will not be in a classroom with the same teacher, the teacher's assessment of the child's looks in the base period will not necessarily be correlated with the child's subsequent performance.

While the ultimate causal role of beauty in affecting educational value-added may be clear, we have not yet examined the proximate mechanisms through which its impact works. In the SECCYD we can get some insight into this by examining how the child's looks alter his/her treatment by the teacher. In each of Waves 5-10 the teacher is asked whether he/she feels close to the student, and whether he/she feels in conflict with the student. Teachers characterize most of their relationships with the student as close and most as basically without conflict—these variables are highly skewed. In modifying the autoregressions,

[14]Applying the same test for biases induced by specification error (Oster, 2019) that we used for the SECCYD shows that the unobservables would need to be nearly as highly correlated with the treatment (looks) as are the observables in the reading scores, and even more strongly correlated with it in the math scores. Both tests suggest that the estimated treatment effects are robust to the possible specification errors arising from unobservables.

we thus create indicators of whether the teacher's closeness (conflict) with the student is in the upper half of the distribution of the measures. We add these indicators sequentially to estimates of the basic autoregression (so that these measures become lagged one period and thus precede the measure of the value-added).

Columns (1) and (4) of the upper panel of Table 11 present re-estimates of the autoregressions in Table 8, but with samples consisting only of Waves 6-11 for comparability to the expanded specifications that include the closeness/conflict indicators. The estimates of the impacts of looks are somewhat smaller than those based on Waves 2-11, but the impact is statistically significant in the equation without controls, and nearly so in that including controls. Columns (2) and (5) add the indicator for teacher-student closeness, while Columns (3) and (6) add the indicator for teacher-student conflict. When the teacher feels close to the student, the student's test score increases more—an effect of about 0.05 standard deviations comparing students in the upper to those in the lower half of this measure. The impact of the teacher feeling in conflict with the students is about the same size but of opposite sign.[15]

The comparisons are of the impacts of beauty on value-added without and then with these measures of the teacher-student relationship. The estimated impacts of looks are smaller when these indicators are included; but the declines are no more than ten percent. A reasonable conclusion is that, while looks affect value-added in schools, their effects do not work through this teacher-reported characterization of the relationship with the student (although potentially an improved – less skewed – measure might mitigate the impacts of differences in looks more strongly).

With the subjective measures of looks in the NCDS produced by teachers, we cannot use teachers' assessments of the child's characteristics to infer the paths through which better looks increase educational value-added. Rather, we use the mother's assessments of their child's behavior, reported in the same wave as the measure of looks and presumably based on observations outside the classroom. To the extent that

[15]Replacing the indicators with the continuous, highly skewed raw measures yields the same qualitative conclusions. Because these measures are highly correlated, including both in the same specification adds little.

these assessments affect student achievement, they work differently from those in the SECCYD, generating effects through greater parental rather than teachers' attention to good-looking children. We use vectors of indicators reflecting mothers' five assessments, each on a scale of "never," "sometimes" and "frequently." These are responses at age 7 to questions about whether the child has difficulty concentrating; whether he/she is upset by new situations; and whether he/she fights with other children. We also use mothers' reports at age 11 about whether the child is miserable or tearful, and whether he/she is squirmy or fidgety.[16]

The first and third columns of the bottom panel of Table 12 re-estimate the models of Table 10 for reading and math test scores at age 16 but using looks measured at age 7 as the base-period. Thus the estimates of the impacts of looks show their effects over the entire period of the child's compulsory schooling. The impacts of the indicators of good and bad looks are quite like those in Table 10, with both having statistically significant impacts on value-added in reading and math. (As before, we include a measure of gender and a vector of indicators of father's social class in the base period.) The second and fourth columns in the lower panel add the five vectors of mother's assessments of the child's behavior. In each case all five vectors have the expected effects on value-added: If the mother reports that the child never exhibits the behavior, the value-added in test scores is higher. Moreover, in most of the cases the vector describing the behavior has a significant impact.

Although mothers' assessments of children behavior are related to improvements in test scores over the nine-year range, their inclusion in the estimates hardly alters the measured impacts of looks. While those do decline, the decreases average around ten percent or less, quite close to the declines observed in the SECCYD. The impacts of children's looks are essentially independent of maternal ratings of child behaviors that might be viewed as detrimental to their achievement. Viewed alternatively, the teacher's assessments of a child's looks are Beautynot greatly affected by mother's perceptions of the child's behavior, or it may be that mothers' ratings do not well reflect children's behaviors in the school and peer contexts closer to teachers' perceptions of them and to their standardized test performance.

[16]Among the three indicators observed at age 7, 69, 71 and 41 percent of mothers report that their child never had this difficulty. On the two reports at age 11, 59 and 60 percent of mothers state their child never exhibited this behavior.

## VII. Inferring How the Economic Returns to Education Arise from Workers' Beauty

The beauty literature (Hamermesh, 2011) has examined the extent to which differences in looks affect economic outcomes, particularly earnings, conditional on large numbers of personal and job characteristics, including educational attainment. It, and the much more massive literature on the returns to education, in one form or another all measure the impact of an additional year of schooling, or an additional degree obtained, on wage rates and/or earnings. As we have shown, however, being better-looking also raises a student's measured achievement. To the extent that greater achievement leads to attaining additional education, part of the effect of education on earnings that has been measured in the immense literature on the returns to education arises indirectly through the effects of beauty.

Ideally, we would like to estimate the following triangular model over individuals in some survey:

(2a) $S_{ct} = F(S_{c,t-1}, \text{Looks}_{c,t-1}, X_{c,t-1})$ , t during childhood, c;

(2b) $ED_{yt} = G(S_{ct}, \text{Looks}_{c,t-1}, X_{ct})$, t during young adulthood, y;

and:

(2c) $\text{Earnings}_{mt} = H(ED_{yt}, \text{Looks}_{c,t-1}, X_{mt})$, t during maturity, m,

where X is a vector of current-period controls. To estimate this model, we need to observe people over much of their lives, at least from the primary grades through their prime earning years. With the SECCYD we cannot do this—we cannot tell whether extra achievement leads to more education (and thus higher earnings); but we can use the results here and extraneous information on the relation between test scores and educational attainment, and educational attainment and earnings, to infer the magnitude of the effect of beauty on earnings through its impact on education. With the NCDS we can estimate this model directly using just this one data set, since the respondents have been followed from age 7 though middle age.[17]

[17]While we do not observe the child's eventual education attainment, the SECCYD does include his/her 9th grade (Wave 11) grade point average. With the same controls as in the expended versions of (1), and including the Wave 10 test score, a one standard-deviation increase in average beauty raises ninth-grade GPA by 0.22 points.

*A. Indirect Effects Inferred though the SECCYD*

Estimates of the impact of looks at each age can be inferred from the pooled autoregressions reported in Table 8. The results of these calculations are shown in the first three columns of Table 12 for all children, then for girls and boys separately. Not surprisingly, given the results for the autoregressions themselves, the impacts of looks for boys are larger than those for girls. On average, these are not huge effects; they are roughly in line with what the education literature suggests is the impact on test scores of moving from a very ineffective teacher to a very effective one.

An alternative is to take each wave separately, using the autoregressions reported in Tables 6 and 7, then concatenate the results up to each wave based on each autoregression and the cumulative effect of looks on test scores up to that point. With separate autoregressions for each wave of the sample and, with the results in Section III suggesting negative conditional impacts of looks on test scores at a few waves, some of the cumulative effects will be negative, since idiosyncratic variations in the small cross-sectional samples generate substantial fluctuations in estimates of the current-period impact. We observe this in Columns (4)-(6) of Table 12. Nonetheless, for the pooled data on girls and boys, the net impact of looks at age 15, shown near the bottom of Column (4), is very close to that based on the pooled autoregression (shown near the bottom of Column (1)).

We performed these calculations in order to use them to infer the indirect impact of looks on earnings that occurs through its effects on educational attainment. Chetty *et al*. (2014a) initially show that a one standard-deviation increase in teacher quality raises test scores by 0.13 standard deviations, somewhat more than the 0.075 (0.146*.53) cumulative effect of looks on scores implied by the estimates in Table 12. Chetty *et al*. (2014b, pp. 2655-56) calculate that such a one standard-deviation increase in tests scores raises earnings by 12 percent. The implied impact of looks on earnings through its effects on educational attainment in the SECCYD is then 0.9 percent (0.075·12 percent), i.e., equivalent to the impact on earnings of 1 extra month of additional schooling (assuming a 12 percent annual return to education).

The estimates of the direct impact of looks on earnings are typically much larger than this, with the equivalent of a one standard-deviation increase in one's position in the distribution of looks (from the median to the 84th percentile) increasing earnings by about 7 percent.[18] The indirect impact of looks on earnings is not tiny—some of the economic returns to education arise from the effect of looks on the treatment of children and its effects on their test scores and hence on their educational attainment. The overwhelming majority of the role of beauty in labor markets results, however, from its direct impacts. In these data perhaps a little more than 10 percent (0.9/[0.9+ 7]) stems from its indirect effect.

*B. Indirect Effects Calculated from the NCDS*

In the NCDS we can estimate the equations (2) individually, separately for each of the test scores, thus creating autoregressions examining changes in reading and math test scores between ages 7 and 16, the longest time span that we observe during childhood. With educational attainment recorded in the NCDS, we use two indicators of educational attainment as the outcomes in (2b): A Levels Plus (averaging 0.41 of the sample respondents), and Degree Plus (averaging 0.27 of the respondents), both measured when the respondent was age 33. In equations describing these outcomes we include both reading and math scores at age 16. We then use these estimates to measure the effect of educational attainment on subsequent earnings observed at ages 33 and 46. In estimating both (2a) and (2b) we hold constant the same variables that we included in the autoregressions of age 11 test scores reported in Table 10. The earnings equations add to these a vector of indicators of the region in the U.K. where the respondent resided at that age.

The estimates of Equation (2a) were shown in Columns (1) and (3) of Table 11, so we do not report them here. We present the estimates of Equation (2b) in Columns (1) and (2) of Table 13. Not surprisingly, higher reading and math scores at age 16 strongly affect the likelihood of having at least A Levels or at least a Degree, with one standard-deviation increases in each raising the probability of at least A levels by 0.29, very large compared to the fraction of respondents with at least A Levels. (Similar large effects are estimated on the probability of attaining a university degree.) The direct effects of looks on the probability

[18] Authors' calculations based on estimates of earnings equations over 8 different data sets from 5 countries.

of obtaining at least a degree are small; on at least having A Levels, they are larger, with the often-observed asymmetric greater response of the outcome to bad looks (e.g., Hamermesh and Biddle, 1994). Remember, however, that looks may indirectly affect educational attainment through their impacts on test scores.

The final four columns of Table 13 present standard log-earnings equations, estimated at ages 33 and 46, and estimated separately with one or the other indicator of educational attainment. Looks have the expected effects on earnings; and differences in educational attainment produce significant impacts on earnings, with obtaining at least a university degree yielding slightly higher returns than obtaining either A levels or a degree.[19]

We can simulate the effect of moving from bad- to good looks (a 2.19 standard-deviation increase) on test scores using the outcomes in Columns (1) and (3) in the lower panel of Table 11. More interesting are the calculations of the indirect effect of looks on educational attainment through test scores, presented along with the estimates of the direct effects in Columns (3) and (4). These demonstrate that at least half of the effect of differences in appearance on educational attainment works indirectly through their effect in raising test scores.

The central results of this subsection are shown in Columns (5)-(8), which take the estimates of the direct effects on earnings at ages 33 and 46 in this sample from Table 13 and calculate the extra impact of looks on earnings arising from their indirect impacts through educational attainment, and hence through test scores. At the bottom the table lists the total effect of differences in looks on earnings. The indirect effects of looks on earning are substantial. Depending on the age when the respondent's earnings are observed and the measure of educational attainment that is used, they comprise between 20 and 50 percent of the total impact.

Using these estimates, a simulated one standard-deviation increase in looks raises earnings in this sample by 6 and 8 percent (total effect/2.19). On a per standard-deviation increase, the indirect effects

[19]To save space we have only presented the earnings equations for ages 33 and 46. The sizes of the estimates for earnings at age 41 are between those shown in the table, while the estimates at age 51 are close to those shown for age 33.

(through test scores, and then through schooling) are between 1.3 and 3.9 percent, and average 2.4 percent in this data set, over twice those inferred from the SECCYD.

Overall the results in this entire section, from two completely independent investigations of the relationship between looks and value-added in education, suggest that a not insubstantial part of the return to schooling arises because better-looking students improve their achievements in school more rapidly than other students, improvements that lead them to obtain a higher level of education. Summarizing, perhaps 10 to 20 percent of the economic returns to schooling arise from the indirect effects of beauty on educational attainment.

**VIII. Conclusions and Implications**

We have engaged in various exercises to examine how looks affect children's intellectual development, measured by the changes in what are mostly objective measures of a child's or adolescent's achievement. One data set, the longitudinal U.S. Study of Early Child Care and Youth Development, followed a sample of over 1000 infants through age 15, collecting information at 11 waves based on a variety of measures of achievement, mostly objective from standardized tests. We replicate the results using the 1958 cohort of the U.K. National Child Development Study, which has followed all children born in the U.K. in a particular week up through middle age, with objective assessments of their achievement at ages 7, 11 and 16. In the SECCYD we employed contemporaries of this cohort to rate their looks based on thin slices of videos taken at each age, taking averages of the normalized ratings of each child's looks at each age. In the NCDS we use the child's teachers' assessments of his/her looks at ages 7 and 11.

Estimating autoregressions describing the change in achievement between waves as affected by these looks measures, and in some specifications by sets of class/income and racial/ethnic indicators, we demonstrate that looks matter—better-looking children show greater improvements in assessments on average. Because students who perform better in primary and secondary school are more likely to obtain additional education, these results imply that some of the labor-market returns to education arise from the indirect effect of looks on educational attainment, perhaps as much as 20 percent of the returns to an additional year of schooling. This indirect effect is in addition to the direct effect of looks on earnings and other economic outcomes. This inference does not mean that schooling is unproductive. Rather, it implies that the benefits of schooling are tilted toward better-looking students, whose good looks lead them to greater achievements in school and to greater educational attainment than their less good-looking contemporaries.

We have explored two plausible mechanisms by which better looks might produce higher achievement—teachers' closeness to and conflict with students, and parents' assessments of children's behavior. Although each was associated in expected ways with looks and gains in achievement, none greatly affected the impact of looks on gains in achievement. Inferring the indirect pathways will require studies designed specifically to consider how lookism might operate from early childhood through adolescence. Developmental scholarship has largely neglected such mechanisms, with research on the beauty bias being primarily concentrated in the social psychological and economics literatures. Promising directions for future studies are suggested by existing studies (e.g., Gordon *et al.*, 2013).

At the higher end of looks, children possess a characteristic that others associate with all kinds of positive attributes (being smarter, friendlier, and more likely to succeed). Not only have such assumptions tied to good looks been shown in experimental studies to alter how others interact with the child, but through a self-fulfilling prophecy they also can lead the child to behave in ways consistent with those expectations (Langlois *et al.*, 2000; Snyder *et al.*, 1977). At the lower end of looks, stigmatizing characteristics negatively affect identity (Goffman, 1963; Major and O'Brien, 2005). Children assessed as being bad-looking may be less often selected for investments and rewards and may be treated in ways that reduce self-esteem and mental health. Together these processes might explain some of the advantages to good looks that we have documented, for instance, with better-looking children being placed in higher reading groups and in accelerated tracks, and with worse-looking children feeling less confident in their abilities when sitting for a standardized test.

Studies are needed that connect what is known from this literature to observational studies tracking the natural unfolding of development and that are specifically focused on looks, rather than the more often studied aspects of status and stigma such as sex/gender, race/ethnicity, and disability. Existing measures of relationships, identities and discrimination can be adapted to measure how others respond to children's looks and how youths internalize those responses, including ratings probing looks-based teasing, avoidance or attraction, and experience-sampling methods capturing how teachers may differentially respond to equally-able students with better-and worse-rated looks. If such measures were embedded into longitudinal studies with the kinds of ratings of attractiveness and standardized achievement used here, the mechanisms generating the robust associations evident here could be better understood.

## REFERENCES


Daniel Aaronson, Lisa Barrow and William Sander, "Teachers and Student Achievement in Chicago Public High Schools," *Journal of Labor Economics*, 25 (Jan. 2007): 95-135.

Ama Abeberese, Todd Kumler and Leigh Linden, "Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines," *Journal of Human Resources*, 49 (Summer 2014): 611-33.

Jason Abrevaya and Daniel Hamermesh, "'Beauty Is the Promise of Happiness'? *European Economic Review*, 64 (2013): 351-68.

Richard Akresh and Ilana Redstone Akresh, "Using Achievement Tests to Measure Language Assimilation and Language Bias among the Children of Immigrants," *Journal of Human Resources*, 46 (Summer 2011): 648-67.

Daniel Benjamin and Jesse Shapiro, "Thin-Slice Forecasts of Gubernatorial Elections," *Review of Economics and Statistics*, 91 (August 2009), pp. 523-36.

Niclas Berggren, Henrik Jordahl and Panu Poutvaara, "The Looks of a Winner: Beauty and Electoral Success," *Journal of Public Economics*, 94 (2010): 8-15.

Daniela del Boca, Chiara Monfardini and Cheti Nicoletti, "Parental and Child Time Investments and the Cognitive Development of Adolescents," *Journal of Labor Economics*, 35 (April 2017): 565-608.

Scott Carrell, Mark Hoekstra and Elira Kuka, "The Long-Run Effects of Disruptive Peers," *American Economic Review*, 108 (Nov. 2018): 3377-415.

Raj Chetty, John Friedman and Jonah Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104 (Sept. 2014): 2593-2632, a.

Raj Chetty, John Friedman and Jonah Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Add and Student Outcomes in Adulthood," *American Economic Review*, 104 (Sept. 2014): 2633-79, b.

Robert Crosnoe, Fred Morrison, Margaret Burchinal, Robert Pianta, Daniel Keating, Sarah Friedman, K.A. Clarke-Stewart, "Instruction, Teacher-Student relations, and Math Achievement Trajectories in Elementary School, *Journal of Educational Psychology*, (2010): 407-17.

Greg J. Duncan, Chantell Dowsett, Amy Claessens, Katherine Magnuson, Aletha Huston, Pamela Klebanov, Linda Pagani, Leon Feinstein, Mimi Engle, Jeanne Brooks-Gunn, Holly Sexton, Kathryn Duckworh and Crista Japel, "School Readiness and Later Achievement," *Developmental Psychology*, (2007): 1428-46.

Susan Dynarski, Daniel Hubbard, Brian Jacob and Silvia Robles, "Estimating the Effects of a Large For-Profit Charter School Operator," NBER Working Paper No. 24428, 2018.

Ralph Waldo Emerson, *The Conduct of Life*. Boston: James R. Osgood, 1871.

Roland Fryer, "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *Quarterly Journal of Economics*, 126 (Nov. 2011): 1755-98.

Jonah Gelbach, "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics*, 34 (April 2016): 509-43.

Erving Goffman, *Stigma*. Englewood Cliffs, NJ: Prentice Hall, 1963.

Rachel Gordon, Robert Crosnoe and Xue Wang, "Physical Attractiveness and the Accumulation of Social and Human Capital in Adolescence and Young Adulthood," *Monographs of the Society for Research in Child Development*, *78* (2013).

Rachel Gordon, Lilla Pivnick, Sarah Moberg and Robert Crosnoe, "Documentation of Appearance Ratings for the Study of Early Child Care and Youth Development (SECCY)" *OpenICPSR*, (2018).

Daniel Hamermesh, *Beauty Pays*. Princeton University Press, 2011.

Daniel Hamermesh and Jeff Biddle, "Beauty and the Labor Market," *American Economic Review*, 84 (Dec. 1994): 1174-94.

Eric Hanushek, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 61 (2): 280–88.

------------------and Steven Rivkin, "Generalizations About Using Value-Added Measures of Teacher Quality," *American Economic Review*, 100 (May 2010): 267-71.

Barry Harper, "Beauty, Stature and the Labour Market: A British Cohort Study," *Oxford Bulletin of Economics and Statistics*, 62 (2000): 771-800.

Elaine Hatfield and Susan Sprecher, *Mirror, Mirror: The Importance of Looks in Everyday Life*. Albany, NY: SUNY Press, 1986.

Linda Jackson, John Hunter and Carole Hodge, "Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review," *Social Psychology Quarterly*, 58 (June 1995): 108-22.

Amy King and Andrew Leigh, "Beautiful Politicians," *Kyklos*, 62 (Nov. 2009): 579-93.

Judith Langlois, Lisa Kalakanis, Adam Rubenstein, Andrea Larson, Monica Hallam and Monic Smoot, "Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review," *Psychological Bulletin*, 126 (May 2000): 390-423.

Brenda Major and Laurie T. O'Brien. "The Social Psychology of Stigma," *Annual Review of Psychology*, 56 (2005): 393-421.

NICHD SECCYD Steering Committee, *Child Care Data Report – 1: Hospital Recruitment Data*. Nashville, TN: Quantitative Systems Laboratory, Peabody College, Vanderbilt University, 1993.

NICHD Early Child Care Research Network, *Child Care and Child Development: Results from NICHD's Study of Early Child Care and Youth Development*. New York: Guilford Press, 2005.

Sonia Oreffice and Climent Quintana-Domeque, "Beauty, Body Size and Wages: Evidence from a Unique Data Set," *Economics and Human Biology*, 22 (Sept. 2016): 22-34.

Emily Oster, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, 37 (April 2019): forthcoming.

Marta Rubio-Codina, Orazio Attanasio, Costas Meghir, Natalia Varela, and Sally Grantham-McGregor, "The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota," *Journal of Human Resources*, 50 (Spring 2015): 465-83.

Joseph Salvia, Robert Algozzine and Joseph Sheare, "Attractiveness and School Achievement," *Journal of School Psychology*, 15 (1977): 60-7.

J. Karl Scholz and Kamil Sicinski, "Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study," *Review of Economics and Statistics*, 97 (March 2015): 14-28.

Mark Snyder, Elizabeth Decker Tanke, and Ellen Berscheid. "Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes," *Journal of Personality and Social Psychology*, 35 (1977): 656-666.

Sean Talamas, Kenneth Mavor and David Perrett, "Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance," *PLoS ONE*, 11(2), Feb. 17, 2016.

U.S. Bureau of the Census, *Money Income of Families, Households and Persons in the United States: 1990, Current Population Reports P60-174*. Washington: GPO, 1991.

Deborah Vandell, Jay Belsky, Margaret Burchinall, Laurence Steinberg, Mathan Vandergrift, NICHD Early Child Care Research Network, "Do Effects of Early Child Care Ezxtend to Age 15 Years? Results from the NICHD Study of Early Child Care and Youth Development," *Child Development*, 81 (May/June 2010): 737-56.

Robert Woodcock and Mary Bonner Johnson, *Woodcock-Johnson Psycho-Educational Battery – Revised*. Allen, TX: DLM, 1989

**Table 1. Number and Percentage of Children with Short Slices of Video at Each Wave**

| Age: Months | Number | Percentage |
|---|---|---|
| 6 | 1270 | 93.1 |
| 15 | 1227 | 90.0 |
| 24 | 1157 | 84.8 |
| 36 | 1164 | 85.3 |
| 54 | 1018 | 74.6 |
| Grade: | | |
| 1 | 987 | 72.4 |
| 3 | 976 | 71.6 |
| 4 | 863 | 63.3 |
| 5 | 948 | 69.5 |
| 6 | 875 | 64.1 |
| Age: | | |
| 15 years | 865 | 63.4 |

*Note*. Percentages are relative to the original sample (N = 1,364).

**Table 2. Percent Distributions of Raw Ratings of Beauty, All Waves (N=141,369)***

| Rating | All |
|---|---|
| Very attractive/very cute | 6.3 |
| Attractive/cute | 31.5 |
| Average | 41.9 |
| Unattractive/not cute | 17.7 |
| Very unattractive/not at all cute | 2.6 |

*The number of videos rated.

**Table 3. Mean and Standard Deviation of Mean and SD Looks (the Means and Standard Deviations within Child/Wave of the Rater/Wave Normalized Raw Ratings)**

| | Mean (SD)[a] | | Standard Deviation of Ratings | |
|---|---|---|---|---|
| **Time [N raters][b]** | **Girls** | **Boys** | **Girls** | **Boys** |
| 6 mos. [35] | 0.035 (0.465) | -0.031* (0.445) | 0.903 | 0.891 |
| 15 mos. [27] | 0.012 (0.478) | -0.006 (0.481) | 0.889 | 0.881 |
| 24 mos. [29] | 0.048 (0.450) | -0.039* (0.437) | 0.881 | 0.896 |
| 36 mos. [29] | 0.059 (0.449) | -0.055* (0.419) | 0.913 | 0.896* |
| 54 mos. [30] | 0.021 (0.441) | -0.018 (0.408) | 0.918 | 0.886* |
| 1st grade [29] | 0.075 (0.509) | -0.067* (0.462) | 0.886 | 0.835* |
| 3rd grade [29] | 0.075 (0.617) | -0.135* (0.512) | 0.842 | 0.784* |
| 4th grade [34] | 0.091 (0.611) | -0.062* (0.553) | 0.826 | 0.792* |
| 5th grade [32] | 0.066 (0.654) | -0.069* (0.557) | 0.822 | 0.782* |
| 6th grade [12] | 0.093 (0.702) | -0.119* (0.576) | 0.778 | 0.749* |
| Age 15 [35] | 0.187 (0.719) | -0.192* (0.652) | 0.683 | 0.703* |
| **All Waves** [45] | 0.069 (0.551) | -0.065* (0.497) | 0.858 | 0.838* |

[a]Standard deviations of mean looks in parentheses.
[b]Total number of raters at each wave. Study youth were rated by at least 10 raters at each wave.
*Different from girls at the 95-percent level of confidence.

**Table 4. Determinants of Mean and SD Looks**[a]

| Dep. Var.: | Mean Looks | | | SD Looks | | |
|---|---|---|---|---|---|---|
| **Ind. Var.** | **All** | **Girls** | **Boys** | **All** | **Girls** | **Boys** |
| Female | 0.133<br>(0.017) | --- | --- | 0.020<br>(0.005) | --- | --- |
| Non-Hispanic White | -0.049<br>(0.028) | -0.063<br>(0.047) | -0.050<br>(0.035) | 0.007<br>(0.011) | 0.013<br>(0.017) | 0.002<br>(0.014) |
| Non-Hispanic Black | -0.213<br>(0.036) | -0.320<br>(0.057) | -0.121<br>(0.044) | 0.060<br>(0.012) | 0.053<br>(0.019) | 0.067<br>(0.017) |
| Non-Hispanic Other | -0.041<br>(0.051) | -0.057<br>(0.078) | -0.048<br>(0.070) | 0.073<br>(0.018) | 0.082<br>(0.028) | 0.066<br>(0.023) |
| Household income at birth: $26,000-$52,000 | 0.025<br>(0.025) | 0.040<br>(0.039) | 0.008<br>(0.032) | -0.001<br>(0.007) | -0.007<br>(0.011) | 0.004<br>(0.010) |
| $52,000-$78,000 | 0.044<br>(0.029) | 0.052<br>(0.049) | 0.033<br>(0.034) | -0.002<br>(0.008) | -0.004<br>(0.014) | -0.001<br>(0.010) |
| $78,000-$275,000 | 0.044<br>(0.032) | 0.069<br>(0.049) | 0.017<br>(0.041) | -.0.008<br>(0.009) | -0.009<br>(0.014) | -0.009<br>(0.012) |
| $\bar{R}^2$ | 0.034 | 0.032 | 0.013 | 0.011 | 0.007 | 0.013 |
| N = | 10,399 | 5,181 | 5,218 | 10,399 | 5,181 | 5,218 |
| N Individuals = | 1,281 | 619 | 662 | 1,281 | 619 | 662 |

[a]Mean and SD looks are the means and standard deviations within child/wave of the rater/wave normalized raw ratings. Standard errors in parentheses. Also included are indicators of mother's education and of the educational attainment of the more educated parent. Standard errors are clustered on each child.

**Table 5. Descriptive Statistics of Outcome Variables[a]**

| Name: | Variable Description | Mean | SD | Range |
|---|---|---|---|---|
| IMPRSO | Observers' Ratings of Mother/Child Behavior, Overall Impression: Wave 1 | 4.22 | 0.69 | [1, 5] |
| MDI | Bayley Mental Development Index: Waves 2, 3 | 108.58 | 14.07 | [63, 150] |
| BKSRCO | Bracken School Readiness Composite: Wave 4 | 14.76 | 9.92 | [0, 50] |
| WJAPSC | Woodcock-Johnson Applied Problems Standard Score: Waves 5, 6, 7, 9, 11 | 102.94 | 15.63 | [41, 153] |
| WASIFC | Wechsler Full Scale IQ: Wave 8 | 106.86 | 14.83 | [59, 149] |
| ASLL | Academic Skills Rating Scale, Language & Literacy Score: Waves 10 (Teacher-rated) | 3.79 | 0.92 | [1, 5] |

[a]The means and standard deviations shown here are for the variable's first use in one of the following text tables as a dependent or lagged dependent variable: IMPRSO--Wave 1; MDI--Wave 2; BKSRCO--Wave 4; WJAPSC--Wave 5; WASIFC--Wave 8; ASLL--Wave 10.

**Table 6. Autoregressions of Normalized Outcome Measures, Waves 2-6[a]**

| **Wave:** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|
| **Age/grade:** | **15 mos.** | **24 mos.** | **36 mos.** | **54 mos** | **Grade 1** |
| **Dep.Var.:** | **MDI** | **MDI** | **BKSRCO** | **WJAPSC** | **WJAPSC** |
| **Lagged Dep. Var.:** | **IMPRSO** | **MDI** | **MDI** | **BKSRCO** | **WJAPSC** |
| Mean looks | 0.168 | 0.048 | 0.031 | 0.073 | -0.011 |
| | (0.066) | (0.050) | (0.058) | (0.059) | (0.061) |
| Lagged dep. var. | 0.070 | 0.416 | 0.365 | 0.409 | 0.594 |
| | (0.033) | (0.025) | (0.030) | (0.030) | (0.030) |
| Female | 0.191 | 0.265 | 0.199 | 0.060 | -0.268 |
| | (0.061) | (0.048) | (0.053) | (0.052) | (0.051) |
| Non-Hispanic White | 0.082 | 0.270 | 0.163 | 0.189 | 0.142 |
| | (0.129) | (0.104) | (0.113) | (0.112) | (0.110) |
| Non-Hispanic Black | -0.496 | -0.195 | -0.075 | -0.209 | -0.070 |
| | (0.153) | (0.123) | (0.135 | (0.134) | (0.132) |
| Non-Hispanic Other | -0.003 | 0.241 | 0.249 | 0.242 | 0.274 |
| | (0.192) | (0.148) | (0.167) | (0.167) | (0.162) |
| Household income at birth: | | | | | |
| $26,000-$52,000 | 0.207 | 0.113 | 0.046 | 0.191 | 0.011 |
| | (0.088) | (0.068) | (0.075) | (0.075) | (0.076) |
| $52,000-$78,000 | 0.307 | 0.164 | 0.268 | 0.140 | -0.047 |
| | (0.100) | (0.078) | (0.086) | (0.085) | (0.085) |
| $78,000-$275,000 | 0.188 | 0.361 | 0.198 | 0.251 | 0.112 |
| | (0.113) | (0.089) | (0.097) | (0.097) | (0.096) |
| $\bar{R}^2$ | 0.093 | 0.432 | 0.350 | 0.390 | 0.449 |
| N Individuals | 1,008 | 1,029 | 1,013 | 927 | 892 |

[a]Standard errors in parentheses. Also included are indicators of mother's education and of the education of the more educated parent.

**Table 7. Autoregressions of Normalized Outcome Measures, Waves 7-11[a]**

| Wave | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|
| Age/grade: | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Age 15 |
| Dep.Var.: | WJAPSC | WASIFC | WJAPSC | ASLL | WJAPSC |
| Lagged Dep. Var: | WJAPSC | WJAPSC | WASIFC | ASLL | ASLL |
| Mean looks | 0.053 | 0.030 | -0.031 | 0.104 | 0.004 |
| | (0.050) | (0.045) | (0.043) | (0.055) | (0.054) |
| Lagged dep. var. | 0.625 | 0.563 | 0.591 | 0.512 | 0.391 |
| | (0.027) | (0.045) | (0.031) | (0.038) | (0.040) |
| Female | -0.027 | 0.082 | -0.142 | 0.023 | -0.184 |
| | (0.048) | (0.054) | (0.051) | (0.067) | (0.070) |
| Non-Hispanic White | -0.002 | 0.014 | 0.182 | -0.062 | 0.032 |
| | (0.104) | (0.105) | (0.104) | (0.142) | (0.139) |
| Non-Hispanic Black | -0.204 | -0.390 | -0.150 | -0.230 | -0.286 |
| | (0.125) | (0.129) | (0.130) | (0.175) | (0.170) |
| Non-Hispanic Other | -0.110 | 0.124 | 0.199 | -0.334 | -0.210 |
| | (0.156) | (0.162) | (0.159) | (0.206) | (0.214) |
| Household income at birth: | | | | | |
| $26,000-$52,000 | 0.075 | 0.055 | 0.061 | 0.276 | 0.014 |
| | (0.071) | (0.078) | (0.077) | (0.096) | (0.103) |
| $52,000-$78,000 | 0.122 | -0.033 | 0.022 | 0.208 | 0.109 |
| | (0.081) | (0.088) | (0.086) | (0.109) | (0.115) |
| $78,000-$275,000 | 0.121 | 0.096 | 0.124 | 0.166 | 0.188 |
| | (0.092) | (0.100) | (0.097) | (0.126) | (0.132) |
| $\bar{R}^2$ | 0.520 | 0.521 | 0.476 | 0.388 | 0.374 |
| N Individuals | 814 | 704 | 723 | 558 | 527 |

[a]Standard errors in parentheses. Also included are indicators of mother's education and of the education of the more educated parent.

**Table 8. Pooled Autoregressions of Normalized Outcomes, SECCYD Waves 2-11***

| | All | Girls | Boys | All | Girls | Boys |
|---|---|---|---|---|---|---|
| Average stzd. beauty | 0.101 | 0.081 | 0.117 | 0.045 | 0.039 | 0.059 |
| | (0.019) | (0.025) | (0.028) | (0.018) | (0.025) | (0.027) |
| Lagged dep. var. | 0.528 | 0.527 | 0.527 | 0.420 | 0.408 | 0.429 |
| | (0.012) | (0.018) | (0.017) | (0.013) | (0.019) | (0.018) |
| Female | | | | 0.047 | --- | --- |
| | | | | (0.024) | | |
| Non-Hispanic White | | | | 0.109 | 0.147 | 0.088 |
| | | | | (0.046) | (0.068) | (0.059) |
| Non-Hispanic Black | | | | -0.270 | -0.155 | -0.366 |
| | | | | (0.056) | (0.079) | (0.076) |
| Non-Hispanic Other | | | | 0.122 | 0.209 | 0.058 |
| | | | | (0.074) | (0.101) | (0.107) |
| Household income at birth: | | | | | | |
| $26,000-$52,000 | | | | 0.104 | 0.132 | 0.078 |
| | | | | (0.035) | (0.051) | (0.047) |
| $52,000-$78,000 | | | | 0.129 | 0.156 | 0.102 |
| | | | | (0.042) | (0.062) | (0.055) |
| $78,000-$275,000 | | | | 0.192 | 0.172 | 0.213 |
| | | | | (0.044) | (0.064) | (0.061) |
| $\bar{R}^2$ | 0.282 | 0.281 | 0.278 | 0.333 | 0.403 | 0.390 |
| N Observations | 8,334 | 4,173 | 4,161 | 8,218 | 4,140 | 4,078 |
| N individuals | 1,237 | 634 | 633 | 1,216 | 596 | 620 |
| % change in beauty effect due to: | | | | | | |
| Race/ethnicity | | | | 61.6 | 69.6 | 53.1 |
| Family income at child's birth | | | | 12.7 | 15.4 | 13.2 |
| Parents' education | | | | 25.6 | 15.0 | 33.7 |

*Standard errors in parentheses, clustered on each child. Also included are indicators of mother's education and the education of the more educated parent. Dep. and lagged dep. vars are:
Wave 3--MDI, MDI; Wave 4--BKSRCO, MDI; Wave 5--WJAPSC, BKSRCO; Wave 6--WJAPSC, WJAPSC; Wave 7—WJAPSC, WJAPSC; Wave 8—WASIFC, WJAPSC; Wave 9—WJAPSC, WASIFC; Wave 10—ASLL, ASLL; Wave 11—WJAPSC, ASLL

**Table 9. Summary Statistics, NCDS, Ages 7, 11 and 16[a]**

| Age | Variable | Mean (S.D.) |
|---|---|---|
| 7 | Good-looking (attractive)* | 0.597 |
| | Average-looking (all others) | 0.316 |
| | Bad-looking (unattractive or abnormal feature) | 0.087 |
| 7 | Southgate group reading test score** | 23.441 (7.057) |
| 7 | Problem arithmetic test score*** | 5.138 (2.471) |
| 11 | Good-looking (attractive)* | 0.581 |
| | Average-looking (all others) | 0.315 |
| | Bad-looking (unattractive or abnormal feature) | 0.104 |
| 11 | Reading comprehension test score** | 16.077 (6.252) |
| 11 | Mathematics test score** | 16.818 (10.333) |
| 16 | Reading comprehension test score** | 25.614 (6.834) |
| 16 | Mathematics test score** | 12.895 (7.000) |

[a]Standard deviation in parentheses below the mean.
*Children described as "underfed" or "scruffy and dirty" are excluded from the analysis.
**Based on means for the sample with test scores at ages 7 and 11.

**Table 10. Effects of Looks on Reading and Math Scores, Changes between Ages 7 and 11, and 11 and 16, NCDS 1958 Cohort[a]**

| | Reading | | Math | |
|---|---|---|---|---|
| | Age 11 | Age 16 | Age 11 | Age 16 |
| Good -looking at t-1 | 0.086 | 0.081 | 0.123 | 0.049 |
| | (0.015) | (0.013) | (0.015) | (0.014) |
| Bad-looking at t-1 | -0.136 | -0.082 | -0.203 | -0.035 |
| | (0.026) | (0.021) | (0.027) | (0.022) |
| Lagged dep. var. | 0.571 | 0.760 | 0.510 | 0.726 |
| | (0.007) | (0.006) | (0.007) | (0.007) |
| Female | -0.142 | -0.046 | 0.007 | -0.173 |
| | (0.014) | (0.012) | 0.014) | 0.013) |
| p-value of F-statistic on class indicators | <0.001 | <0.001 | <0.001 | <0.001 |
| $\bar{R}^2$ | 0.431 | 0.639 | 0.392 | 0.606 |
| N Individuals | 12,683 | 10,307 | 12,683 | 10,307 |

[a]Standard errors in parentheses.

**Table 11**. **Sources of the Beauty Effect on Value-added, SECCYD and NCDS 1958 Cohort**[a]

**SECCYD Waves 6-11**

| | No controls | | | With controls* | | |
|---|---|---|---|---|---|---|
| **Ind. Var.** | | | | | | |
| Average stdzd. beauty | 0.069 (0.020) | 0.068 (0.020) | 0.063 (0.020) | 0.030 (0.020) | 0.029 (0.020) | 0.028 (0.020) |
| Lagged dep. var. | 0.629 (0.015) | 0.627 (0.015) | 0.623 (0.015) | 0.532 (0.018) | 0.532 (0.018) | 0.531 (0.018) |
| Teacher feels close to student | | 0.060 (0.022) | | | 0.042 (0.022) | |
| Teacher feels in conflict with student | | | -0.083 (0.023) | | | -0.032 (0.023) |
| $\bar{R}^2$ | 0.393 | 0.421 | 0.395 | 0.394 | 0.421 | 0.421 |
| N = | 4,300 | 4,241 | 4,241 | 4,300 | 4,241 | 4,241 |

**NCDS 1958**** Test Score Age 16

| | Reading | | Math | |
|---|---|---|---|---|
| Good Looks Age 7 | 0.078 (0.019) | 0.073 (0.019) | 0.097 (0.020) | 0.091 (0.020) |
| Bad Looks Age 7 | -0.146 (0.033) | -0.130 (0.033) | -0.19 (0.036) | -0.178 (0.036) |
| Test Score Age 7 | 0.567 (0.009) | 0.557 (0.009) | 0.415 (0.009) | 0.403 (0.009) |
| Difficulty concentrating Age 7 | | b | | b |
| Upset by new situations Age 7 | | c | | c |
| Fights other kids Age 7 | | b | | b |
| Miserable or tearful Age 11 | | b | | b |
| Squirmy, fidgety Age 11 | | c | | b |
| $\bar{R}^2$ | 0.412 | 0.415 | 0.316 | 0.33 |
| N | 8,143 | 8,143 | 8,143 | 8,143 |

[a]Standard errors in parentheses below coefficient estimates. Standard errors in the SECCYD are clustered on individuals.

[b]Vector of indicators statistically significant at the 5-percent level of confidence.

[c]Vector of indicators not statistically significant at the 5-percent level of confidence.

*Also included are all the controls included in the estimates in Table 8.
**Also included are an indicator of gender and a vector of indicators of the father's social class when the child was 7.

**Table 12. Cumulative Effects of a One Standard Deviation Increase in the Average Beauty Rating on Development Outcomes (in Standard Deviations of the Standardized Outcome Measures), SECCYD**

| Based on: | Pooled Regressions | | | Cumulated Individual Regressions | | |
|---|---|---|---|---|---|---|
| Age/Grade | All | Girls | Boys | All | Girls | Boys |
| 15 mos. | 0.085 | 0.072 | 0.119 | 0.317 | 0.244 | 0.422 |
| 24 mos. | 0.120 | 0.101 | 0.170 | 0.217 | 0.180 | 0.238 |
| 36 mos. | 0.135 | 0.113 | 0.192 | 0.138 | 0.337 | -0.170 |
| 54 mos. | 0.142 | 0.118 | 0.201 | 0.194 | 0.230 | 0.102 |
| 1st grade | 0.144 | 0.120 | 0.205 | 0.096 | -0.054 | 0.214 |
| 3rd grade | 0.146 | 0.120 | 0.207 | 0.150 | 0.128 | 0.013 |
| 4th grade | 0.146 | 0.121 | 0.208 | 0.141 | 0.195 | -0.023 |
| 5th grade | 0.146 | 0.121 | 0.208 | 0.024 | -0.017 | -0.055 |
| 6th grade | 0.146 | 0.121 | 0.208 | 0.331 | 0.197 | 0.515 |
| Age 15 | 0.146 | 0.121 | 0.208 | 0.137 | -0.013 | 0.461 |
| SD standardized beauty | 0.529 | 0.551 | 0.497 | 0.529 | 0.551 | 0.497 |

**Table 13. Educational Attainment and Earnings, Equations (2a) and (2b), NCDS 1958 Cohort[a]**

| | Dep. Var.: | | | | | |
|---|---|---|---|---|---|---|
| **Ind. Var.** | **A Level Plus*** | **Uni Degree Plus*** | **Age 33**** | **ln(Earnings) Age 33**** | **Age 46**** | **Age 46**** |
| Good Looks Age 7 | 0.0032 (0.0108) | 0.0013 (0.0102) | 0.0349 (0.0161) | 0.0359 (0.0161) | 0.0660 (0.0275) | 0.0747 (0.0275) |
| Bad Looks Age 7 | -0.0600 (0.0197) | -0.0043 (0.0187) | -0.0518 (0.0290) | -0.0770 (0.0290) | -0.0261 (0.0514) | -0.0507 (0.0514) |
| Reading Score Age 16 | 0.0925 (0.0069) | 0.0609 (0.0065) | | | | |
| Math Score Age 16 | 0.1869 (0.0066) | 0.1687 (0.0094) | | | | |
| A Level Plus | | | 0.3799 (0.0158) | | 0.6104 (0.0265) | |
| Degree Plus | | | | 0.3979 (0.0169) | | 0.6398 (0.0280) |
| $\bar{R}^2$ | 0.321 | 0.260 | 0.387 | 0.385 | 0.209 | 0.208 |
| N = | 6,996 | 6,996 | 7,411 | 7,411 | 5,202 | 5,202 |
| Dep. Var. Mean (SE) | 0.412 | 0.269 | £218.39 18.44 | £218.39 18.44 | £380.65 13.27 | £380.65 13.27 |

[a]Standard errors in parentheses below coefficient estimates.

*Also included are an indicator for gender and a vector of indicators of the person's father's social class when the person was age 7.

**Includes all the variables in the other equations plus a vector of indicators of region of residence when the earnings were observed.

**Table 14. Effects of Looks on Test Scores, Educational Attainment and Earnings, NCDS 1958 Cohort Comparison of Bad Looks at Age 7 to Good Looks at Age 7**

| **Outcome:** | **Test Score** | | **Education** | | **ln(Earnings)** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Reading** | **Math** | **A Level Plus** | **Degree Plus** | **Age 33** | **Age 33** | **Age 46** | **Age 46** |
| **Direct Effect:** | 0.217 | 0.297 | 0.063 | 0.006 | 0.087 | 0.113 | 0.092 | 0.125 |
| **Indirect Effect:** | | | | | | | | |
| Through Scores | | | 0.076 | 0.063 | | | | |
| Through Education | | | | | 0.052 | 0.027 | 0.085 | 0.044 |
| **Total Effect** | | | 0.139 | 0.069 | 0.139 | 0.140 | 0.177 | 0.169 |

**Appendix Table A1. Terminology and Calculations for SECCYD Appearance Ratings**

| | |
|---|---|
| Raw ratings | 10 or more undergraduate raters rate each SECCYD youth at each wave |
| Rater/wave normalized ratings (i.e., *normalized ratings*) | Raw ratings are normalized to adjust for rater effects within each wave by subtracting the rater's average and dividing by the rater's standard deviation of ratings for that wave. |
| Youth/wave mean of normalized ratings (i.e., *mean looks*) | The mean of the 10 or more rater/wave normalized ratings of each SECCYD youth is calculated at each wave. |
| Youth/wave SD of normalized ratings (i.e., *SD looks*) | The standard deviation of the 10 or more rater/wave normalized ratings of each SECCYD youth is calculated at each wave. |

**Appendix Table A2. Percentage Distributions, Control Variables, SECCYD, All Observations**

**Variable:**

| | | | |
|---|---|---|---|
| Female | 49.5 | Mother's Education: | |
| Non-Hispanic White | 77.5 | HS or less | 31.2 |
| Non-Hispanic Black | 11.9 | Some college | 33.4 |
| Non-Hispanic Other | 4.6 | Bachelors | 20.8 |
| Hispanic | 6.0 | > Bachelors | 14.6 |
| Household Income at Birth: | | Higher Educated Parent's Education: | |
| <$26,000 | 24.6 | HS or less | 22.7 |
| $26,000-$52,000 | 34.2 | Some college | 33.1 |
| $52,000-$78,000 | 23.1 | Bachelors | 20.9 |
| $78,000-$275,000 | 18.1 | >Bachelors | 23.3 |

## Appendix Table A3. Alternative Specifications, SECCYD[a]

| Ind. Var. | (1a) | (1b) | (2a) | (2b) |
|---|---|---|---|---|
| Mean looks | 0.0799 | 0.0314 | 0.0704 | 0.0219 |
| | (0.0184) | (0.0179) | (0.0186) | (0.0185) |
| Lagged dep. var. | 0.5861 | 0.4740 | 06099 | 0.5016 |
| | (0.0121) | (0.0137) | (0.0121) | (0.0138) |
| $\bar{R}^2$ | 0.345 | 0.394 | 0.375 | 0.421 |
| N = | 7,312 | 7,210 | 6,804 | 6,706 |

| | (3a) | (3b) | (4a) | (5b) |
|---|---|---|---|---|
| Mean looks | 0.0733 | 0.0182 | 0.0489 | 0.0009 |
| | (0.0192) | (0.0184) | (0.0139) | (0.0184) |
| Lagged dep. var. | 0.5071 | 0.3878 | 0.5618 | 0.4352 |
| | (0.0139) | (0.0144) | (0.0139) | (0.0152) |
| $\bar{R}^2$ | 0.258 | 0.323 | 0.315 | 0.380 |
| N = | 8,330 | 8,223 | 7,317 | 7,215 |

[a]The Column (a) in each pair excludes the controls used in Table 8, the Column (b) includes them.
(1) Same as Table 8 without Wave 2.
(2) Same as Table 8 without Waves 2 or 10.
(3) Same as Table 8 using Woodcock-Johnson Picture-Vocabulary Score in Waves 5, 6, 7, 9 and 11.
(4) Same as Table 8 using Woodcock-Johnson Picture-Vocabulary Score in Waves 5, 6, 7, 9 and 11, without Wave 2.