Testing for Racial Differences in the Mental Ability of Young Children*

Roland G. Fryer, Jr. and Steven D. Levitt

Original Submission: March 2006
Current Revision: February 2007

* We would like to thank George Akerlof, Lawrence Katz, Kevin Lang, John List, Robert Moffitt, Kevin Murphy, Andrei Shleifer, Lawrence H. Summers, and three anonymous referees for helpful comments and discussions. Katherine Barghaus, Sheldon Bond, Steve Cicala, Walter Lamberson, Ethan Lieber, and Marina Niessner provided truly outstanding research assistance. Financial support for this study was provided by the National Science Foundation, the Milton Fund at Harvard University, the Sherman Shapiro Research Fund, and the Alphonse Fletcher Sr. Fellowship. Mailing addresses: Roland G. Fryer, Jr., Department of Economics, 1805 Cambridge Street, Harvard University, Cambridge, MA 02138; Steven Levitt, Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637. e-mail: rfryer@fas.harvard.edu; slevitt@uchicago.edu. The usual caveat applies.

Abstract

On tests of intelligence, Blacks systematically score worse than Whites. Some have argued that genetic differences across races account for the gap. Using a newly available nationally representative data set that includes a test of mental function for children aged eight to twelve months, we find only minor racial differences in test outcomes (0.06 standard deviation units in the raw data) between Blacks and Whites that disappear with the inclusion of a limited set of controls. Relative to Whites, children of all other races lose ground by age two. We confirm similar patterns in another large, but not nationally representative data set. A calibration exercise demonstrates that the observed patterns are broadly consistent with large racial differences in environmental factors that grow in importance as children age. Our findings are not consistent with the simplest models of large genetic differences across races in intelligence, although we cannot rule out the possibility that intelligence has multiple dimensions and racial differences are present only in those dimensions that emerge later in life.

Blacks in the United States have consistently scored worse than Whites on tests of IQ and academic achievement (Shuey, 1958; Jensen, 1973, 1998; McGurk et al., 1982; Hernstein and Murray, 1994). Among teenagers and adults, the Black-White test score gap is as much as one standard deviation in magnitude. Large racial gaps in test scores have been found in children as young as two years old (Scott and Sinclair, 1997), and the full racial gap observed later in life is present by age three (Rushton and Jensen, 2005). Even after accounting for a host of demographic and socio-economic factors such as parental income, education, occupation, home environment, birth weight, region, and urbanicity, a substantial Black-White test score gap generally remains.[1] Asians, on the other hand, tend to have systematically higher mean test scores than those of other races (Campbell et al., 1966; Burkett et al., 1995; Rushton, 1995; Fryer and Levitt, 2004).

Some scholars have argued that the combination of high heritability of intelligence (typically above .6, see, for instance, Neisser, 1996 or DeFries, McGuffin, McClearn, and Plomin, 2000) and persistent racial gaps in test scores is evidence of genetic differences across races (Jensen, 1973, 1998; Rushton and Jensen, 2005). As Nisbett (1998) and Phillips et al. (1998) argue, however, the fact that Blacks, Whites, and Asians grow up in systematically different environments makes it difficult to draw strong causal genetically based conclusions.

---

[1] See, for instance, (Campbell et al., 1966; Baughman and Dahlstrom, 1968 Scarr, 1981; Kaufman and Kaufman, 1983; Naglieri, 1986; Bracken, Insko and Sabers, 1987; Humphreys, 1988; Krohn and Lamp, 1989; Brooks-Gunn, 1993a; Brooks-Gunn, Duncan and Kelbanov, 1994, 1996; Phillips et al., 1998; Phillips, 2000; Coley, 2002). A notable exception to this pattern emerges in the Early Childhood Longitudinal Study Kindergarten (ECLS-K) sample of children who entered kindergarten in 1998. This nationally representative sample differs from prior data in that the raw Black-White test score gap at the time children enter kindergarten is substantially smaller than in most prior studies (0.64 standard deviations in math, and 0.41 standard deviations in reading), and the inclusion of a small number of demographic and socio-economic controls erases the gap (Fryer and Levitt, 2004). The Black-White test score gap grows substantially through the first four years of school and by the end of third grade these controls no longer account for the differences in test scores across races (Fryer and Levitt, forthcoming).

Data on mental function in the first year of life represents a potentially important piece of evidence to inform this debate.[2] To the extent that environmental factors play a smaller role at early ages (or alternatively, researchers are better able to measure and control for the environmental factors affecting infants), the presence of an early racial gap in test scores would bolster the argument in favor of a genetic basis for racial differences. On the other hand, an absence of racial differences in mental abilities among children age eight to twelve months substantially complicates (but does not rule out) a genetic basis for a racial IQ gap. To the extent that some aspects of adult intelligence only emerge at later stages of development, or that these aspects go unmeasured in the early test of mental function, the genetic story cannot be definitively rejected with these data.

Because of data limitations, prior research has not been able to directly address this question (Rushton and Jensen, 2005). Studies measuring cognitive abilities of young children have been small-scale, rare, and based on convenience samples that are not drawn with the goal of being nationally representative. For instance, many of the samples involve a limited number of babies born in a particular hospital or metropolitan area (Gravem, Ireton and Thwing, 1970; Wilson 1983), preterm infants (Rose and Wallace, 1985), those with birth weights less than 1,500 grams (Dezoete, MacArthur and Tuck, 2003), or children with rare medical conditions (McGrath et al., 2004). The Early Childhood Longitudinal Study Birth Cohort (ECLS-B) is the first large, nationally representative sample with measures of mental functioning (a shortened version of the Bayley Scale of Infant Development (BSID)) for children aged one and under.

[2] The usefulness of early tests of mental function depends critically on their being predictive of future test scores. Measures of cognitive ability at one year of age (similar to those used in this study) have been found to be strong predictors of test scores later in life by some researchers (Bradley, Caldwell and Elardo, 1975; Wilson, 1983; DiLalla et al., 1990; McGarth, 2004), although other studies report a weaker relationship (Honzik, 1983; Rose and Wallace, 1985; Kopp and McCall, 1990; Anderson, Sommerfelt and Sonnander, 1998). This evidence is discussed in greater detail later in the paper.

Analyzing these data, we find extremely small racial differences in mental functioning of children aged eight to twelve months. With no controls at all, the mean White infant outscores the mean Black infant by .055 standard deviation units – only a tiny fraction of the one standard deviation racial gap typically observed at older ages. The raw scores for Blacks are indistinguishable from Hispanics and Asians, who also slightly underperform Whites. Adding interviewer fixed-effects and controls for the child's age, gender, socio-economic status, home environment, and prenatal circumstances further compresses the observed racial differences. With these covariates, we cannot reject equality in test scores across any of the racial/ethnic groups examined. We obtain similar results for babies under the age of one using a second large, but not nationally representative data set, the U.S. Collaborative Perinatal Project (CPP).

Consistent with Dickens and Flynn (2006), we observe a growing test score gap with age in both of these data sets. By age two (the latest available wave in ECLS), Whites on average outperform children of other races by .3-.4 standard deviations in the raw data, and by .2-.3 standard deviations with the inclusion of extensive controls. In the CPP data, by age four the raw test score gap between Whites and both Hispanics and Blacks is more than .7 standard deviations in the raw data. With controls included, the gap at age four is roughly .3 standard deviations.

When we calibrate these results to a simple model in which test scores are affected by both genes and environment, the characterization that best matches the observed data is one in which there are small mean racial differences in intelligence, environmental factors become increasingly important determinants of test scores as children age, and there are large mean racial differences in environment. To match the correlations in the data using only environmental factors requires Blacks and Hispanics to have an environmental deficit of .60 standard deviations at age 2 and 1.28 standard deviations at ages 4 to 7. Fryer and Levitt (2004) estimate that the gap between Blacks and Whites in socioeconomic status is .75 standard deviations. For number of children's books in the home, the gap is .83 standard deviations, and for female-headed households the gap is 1.3 standard deviations. For variables related to neighborhood safety, even larger gaps are present. Thus, it appears that observed environmental differences can go a long ways towards explaining the observed trajectory of racial test score gaps.

These findings pose a substantial challenge to the simplest, most direct, and most often articulated genetic stories regarding racial differences in mental function. They do not, however, preclude systematic genetic differences across races as an explanation for later observed test score gaps if, for instance, racial gaps are concentrated in higher-order thinking which only emerges later in life. The late emergence of racial differences in test scores is also consistent with the existence of a gene-environment interaction such as argued by Dickens and Flynn (2001). In their model, a positive feedback loop exists between genes and environment, which over time serves to magnify small initial differences when genetics and environmental circumstances are positively correlated.

The remainder of the paper is structured as follows. Section II describes the data. Section III presents the basic empirical estimates. Section IV develops a simple model of genes and environment and explores the extent to which particular characterizations of the model are able to reproduce the observed patterns in the data. Section V concludes.

II. The Data

The primary data we analyze, the Early Childhood Longitudinal Survey Birth Cohort (ECLS-B), is a nationally representative sample of over 10,000 children born in

2001. The first wave of data collection was performed when most of the children were between eight and twelve months of age.[3] The second wave interviewed the same set of children around their second birthday. The data set includes an extensive array of information from parent surveys, interviewer observation of parent-child interactions, and mental and motor proficiency tests. Further details on the study design and data collection methods are available at the ECLS website (http://nces.ed.gov/ecls).

To measure mental proficiency, the ECLS-B uses an abbreviated version of the BSID known as the Bayley Short Form–Research Edition (BSF-R), which was designed to measure the development of children early in life in 5 broad areas: exploring objects (e.g., reaching for and holding objects), exploring objects with a purpose (e.g., trying to determine what makes the ringing sound in a bell), babbling expressively, early problem solving (e.g., when a toy is out of reach, using another object as a tool to retrieve the toy), and naming objects.[4] A child's score is reported as a proficiency level, ranging from 0 to 1 on each of the five sections. These five proficiency scores have also been combined into an overall measure of cognitive ability using standard scale units. Most of our analysis focuses on this overall metric. The test is administered by a trained interviewer and takes twenty-five to thirty-five minutes to complete.

Because this particular test instrument is newly designed for ECLS-B, there is little direct evidence regarding the correlation between performance on this precise test and outcomes later in life. There is, however, a substantial literature of longitudinal studies relating BSID test scores of children aged eight to twelve months to tests of intelligence later in life. Figure 1 provides a graphical representation of the correlations

[3] These children will eventually be resurveyed four times between the present and first grade. At the present time, only the first two waves of data are available.
[4] See Nord, et. al. (2006) for further details.

observed in this prior literature, weighting the estimates of the various papers by the number of subjects included in each study.[5] The horizontal axis is the age in years that the subsequent tests were administered and the vertical axis contains the average correlation between the BSID score and another measure of intelligence. The figure shows that the average correlation between BSID and future IQ scores starts very high and decreases as children age, stabilizing with an average correlation around 0.3 at approximately five years of age.[6] For purposes of comparison, when older children are given achievement scores three years apart, the correlation between scores is on the order of 0.6 (Cruse et al., 1996).[7]

A correlation of .3 between the BSID and future measures of IQ means that the BSID score explains only nine percent of the variation in future test scores for a particular individual. Even though one can explain relatively little of the within-person variation over time in test scores, one would still expect to observe large differences in mean test scores by race on the BSID given the prior evidence. Past research (Wilson and Matheny 1983, Yeates et al. 1983) suggests a correlation between maternal IQ (assessed using the Wechsler Adult Intelligence Scale at the time of interview) and the BSID of twelve month old children of .30. For purposes of comparison, the correlation between maternal IQ (using the same metric) and the well-known Stanford-Binet at three years old is .39. Given the observed one standard deviation in maternal IQ between Whites and

---

[5] Appendix Table 1 list the full set of papers and reported correlations on which Figure 1 is based.

[6] The correlation between the BSR-F in the first and second waves is .26, slightly lower than past results for the BSID.

[7] Other infant tests, such as the Cattell Infant Intelligence Scale, Gesell Development Schedules, or Brunet-Lezine Development Scales, have a substantially lower correlation with later achievement than does the BSID (Escalona and Moriarty, 1961; Karlberg et al., 1968; Birns and Golden, 1972).

Blacks, a correlation of .30 between child's BSID and mother's IQ would imply expected mean differences between White and Black one-year olds of .30 standard deviations.[8]

The ECLS-B sample includes observations on 10,688 children. For 556 of these individuals, no mental ability test was performed in the first wave. Test scores are missing for an additional 1,326 children in the second wave. All subjects with missing test scores are dropped from the analysis. This is the only exclusion we make from the sample.[9] Throughout the analysis, the results we report are weighted to be nationally representative using sampling weights included in the data set.[10]

Table 1 provides summary statistics by major racial/ethnic group in ECLS-B. For all of the background characteristics, we use the values reported in the first wave of data. The mutually exclusive and collectively exhaustive racial/ethnic categories we report are: non-Hispanic Whites (which we simply deem "White"), non-Hispanic Blacks ("Black"), Hispanics, Asians/Pacific Islander ("Asian"), and other race (which combines children characterized as Native American, mixed race, or other race). The top panel of the table reports means and standard deviations by race on the overall measure of mental ability. For ease of interpretation of the regressions, the overall test score has been normalized to have a mean of zero and a standard deviation of one for the sample as a whole. Whites score .013 standard deviations better than the sample mean on the overall mental measure at age 9 months and .176 standard deviations above average around age two.

[8] Indeed, given the results cited in Rushton and Jensen (2005) regarding one standard deviation racial gaps in IQ tests administered at age 3 and the stability in Black-White differences over time, one might expect to observe racial differences in the BSID of close to one standard deviation.
[9] In cases where there are missing values for another of the covariates, we set these missing observations equal to zero and add an indicator variable to the specification equal to one if the observation is missing and equal to zero otherwise. We obtain similar results for the first wave when we include all children with an initial test score, including those who subsequently are not tested in the second wave.
[10] A comparison of the ECLS-B sample characteristics with known national samples such as the US Census and the Center for Disease Control's Vital Statistics confirms that the sample characteristics closely match the national averages.

The next panel of the table presents basic demographic characteristics, which are generally similar across groups. Age at testing is approximately equal across races. As would be expected, Blacks are over represented in the South and underrepresented in the West. Asians and Hispanics are seen in greatest numbers in the West. The fraction of girls and boys are similar across all racial groups.

The third panel has variables capturing the home environment, including socio-economic status (SES) quintiles, number of siblings, family structure, mother's age, and an interviewer rating of the effectiveness of the "parent as a teacher" based on observation of parent-child interactions in a structured problem-solving environment.[11] The socio-economic status measure is constructed by ECLS and includes parental income, occupation, and education. Whites and Asians are concentrated in the higher ranges of the SES distribution; Blacks, Hispanics, and the "other race" category have below average SES. Roughly 90 percent of White and Asian infants are living in households with two biological parents compared to only 41 percent among Blacks. On average, children in our sample have approximately one sibling; Asian children have slightly fewer and Black kids slightly more. White and Asian mothers tend to be older. White parents fare better than the other racial groups on the interviewer evaluation of "parent as teacher" effectiveness.

The final panel of Table 1 presents statistics on the prenatal circumstances of the children. Extremely low birth weight (<1,500 grams) and premature birth are most common among Blacks and least frequent for Whites and Asians. Twins and higher

[11] The "parent as a teacher evaluation" is based on the Nursing Child Assessment Teaching Scale (NCATS). The NCATS scale is composed of 73 binary (yes/no) items that are scored by trained observers. The NCATS coding system has two main scales: the parent scale, which has 50 items, and the child scale with 23 items. The parent scale of the NCATS focuses on the caregiver's use of a "teaching loop," which consists of four components: (1) getting the child's attention, and setting up the expectations for what is about to be done; (2) instruction giving; (3) letting the child respond to teaching; and (4) feedback on the child's attempts to do the task (Nord, 2004). The total parent score can range from 0 to 50 (Yes=1, No=0).

order births are much more frequent among Whites, due primarily to the greater use of infertility techniques such as in vitro fertilization (Hamilton and McManus, 2004).

The second data set used in our analysis is from the Collaborative Perinatal Project (CPP). The data consists of over 31,000 women who gave birth in twelve medical centers between 1959 and 1965. All medical centers were in urban areas; six in the Northeast, four in the South, one in the West and one in the north-central region of the U.S. Some institutions selected all eligible women, while others took a random sample.[12] The socioeconomic and ethnic composition of the participants is representative of the population qualifying for medical care at the participating institutions. These women were re-surveyed when their children were eight months, four years, and seven years old. Follow-up rates were remarkably high: 85% at eight months, 75% at four years and 79% at seven years. We only include students in our analysis that had score results for all three tests.[13]

Table 1B reports summary statistics for the CPP data. The top panel of the table corresponds to the test of mental function, reported separately at ages eight months, four years, and seven years. At eight months, the full Bayley Scale of Infant Development was administered. When children were re-surveyed around their fourth and seventh birthday, the Stanford-Binet and Wechsler Intelligence Test were administered, respectively. Despite the fact that the CPP was carried out three decades earlier than the ECLS,the racial differences in test scores at age 8 months are similar in magnitude to nine-month ECLS scores. By ages four and seven, the raw racial gaps are substantial. The remainder of the table presents summary statistics for demographics, measures of

[12] Detailed information on the selection methods and sampling frame from each institution can be found in Niswander and Gordon (1972).

[13] Analyzing each wave of the data's test scores, not requiring that a student have all three scores, yields similar results.

home environment, and prenatal factors. In all cases, we use the values collected in the initial survey for these background characteristics. The available home environment variables include information on parental education, occupational status, family income, and family structure.[14] These summary statistics clearly demonstrate the non-representative composition of the sample (e.g., roughly half the sample is black and less than ten percent of the parents have a college degree).

III. Racial Differences in the Mental Ability of Young Children

Figures 2 and 3 plot the density of mental test scores by race at various ages in the ECLS and CPP data sets respectively. In Figure 2, the test score distributions on the BSF-R at age nine months for children of different races are essentially indistinguishable visually. By age two, the White distribution has demonstrably shifted to the right. Figure 3 shows a similar pattern using the CPP data. At age eight months, all races look similar. By age four, Whites are far ahead of Blacks and Hispanics. Figures 2 and 3 make the main point of the analysis, namely that the commonly observed racial gaps in test scores emerge only later in life. In the regression analysis that follows, we control for a wide range of other factors, but the basic conclusions suggested by Figures 2 and 3 are unchanged.

Our regression approach involves weighted least squares estimation of equations taking the general form:

$$MENTAL\ TEST_i = \sum_R \beta_R R_i + \Gamma X_i + \varepsilon_i \qquad (1)$$

[14] It must be noted, however, that there is a great deal of missing data on covariates in CPP; in some cases more than half of the sample has missing values. We include indicator variables for missing values for each covariate in the analysis that follows.

where *i* indexes individuals and *r* corresponds to the racial group to which an individual belongs. The vector *X* captures a wide range of possible control variables, and ε is an error term. Also included in all specifications are interviewer-fixed effects, which adjust for any mean differences in scoring of the test across interviewers.[15]

The basic results for the coefficients on the race variables in the first wave of ECLS are presented in Table 2. The omitted race category is "Whites," so the other race coefficients are relative to that omitted group. Each column reflects a different regression. The first column has no controls. As one moves from left to right in Table 2, the set of covariates is progressively expanded. The coefficients on the other covariates are not shown in the table, but full estimation results can be found in Appendix Table 2. The second column adds interviewer-fixed effects. Column (3) adds controls for age at which the test is administered, the gender of the child, and region. Because the age at which the test is taken is such an important determinant of test performance, we include separate indicators for months of age in our specification. The R-squared in the regression jumps substantially with the inclusion of the age and gender controls (from .14 in the first column to .71 in the second column); which is mainly due to the age controls. A child taking the test at age eleven months is predicted to perform .64 standard deviations better than the same child when tested at age nine months, according to our estimates. Girls outscore boys by .05 standard deviations on average, a difference that persists throughout all our specifications. Including these controls improves the

[15] In ECLS, each of the 13 regions was staffed by one field supervisor and between 14 and 19 interviewers, for a total of 256 field staff (243 interviewers), who conducted an average of 42 child assessments each. The number of interviews per interviewer range from 1 to 156. Almost all interviewers assessed children from different races (James et al., 2004). There are 184 interviewers in CPP for 8 month olds, 305 for four year olds and 217 for seven year olds. In the CPP, there are many interviewers for whom virtually all of the children assessed were of the same race.

performance of Whites versus other races because Whites are slightly younger on average when tested.

Column (4) adds indicators for the family's SES quintile. The SES variable (not shown in the main table) enters with the expected sign, i.e. higher SES children score higher on the test, but the magnitude of the effects are small (a top quintile SES child outscores a bottom quintile child by .08 standard deviations). Inclusion of the SES variable shrinks the coefficient on Black and "other race."

Adding a range of other controls for a child's home environment (family structure, mother's age, number of siblings, and parent as teacher measure) shrinks the gap between Whites and each of the other races. None of the race coefficients are statistically significant.

The final column adds controls for a range of prenatal condition variables (birth weight, premature birth, and multiple births). The coefficient on Black actually becomes slightly positive. All the point estimates are substantively small: the largest gap is only .015 standard deviation units.[16]

Table 3 presents results for the second wave of ECLS and for the three ages of test data in CPP. For each age and data set, we report two specifications. The odd numbered columns simply report raw differences in test scores (parallel to column 1 of Table 2). The even columns show results with the most exhaustive set of controls (mirroring the final column of Table 2). In all cases, the omitted race category is Whites and all estimates are relative to that group.

[16] We thoroughly explored the sensitivity of these conclusions to a wide range of alternative specifications , sub-groups of the population, and particular components of the test in an earlier version of this paper (Table 3, Fryer and Levitt 2006). The results prove quite robust. Of particular importance is that controlling for motor skills (a dimension on which Blacks outperform Whites at early ages and which might influence measurement of mental function), did not alter the results appreciably.

For purposes of comparison, we reproduce the results for 9 month olds in ECLS in the first two columns of Table 3. The next two columns correspond to these same children tested around age two. Raw gaps of almost .4 standard deviations between Blacks and Whites are present on the test of mental function at age 2. Even after including extensive controls, a Black-White gap of more than .2 standard deviations remains. Hispanics look similar to Blacks. Asians lag Whites, but by a smaller magnitude.

The CPP data shows similar patterns. Columns 5 and 6 of Table 3 report results from CPP at age 8 months. These results for Blacks are quite similar to the ECLS estimates at age 9 months in columns 1 and 2 of the table: raw differences of less than .1 standard deviations that shrink and become statistically insignificant with the inclusion of controls. Hispanics actually outperform whites in the raw data in the early wave of CPP (although this appears to be solely an artifact of systematic differences across the set of interviewers who tested Hispanic children relative to those of other races), but do worse with controls. By age four, however, a large test score gap has emerged for Blacks, Hispanics, and "other race." In the raw data, Blacks lag Whites by almost .8 standard deviations and Hispanics fare even worse. The inclusion of controls reduces the gap to roughly .3 standard deviations for Blacks and .5 standard deviations for Hispanics. The age seven results are generally similar to those at age four.

V. Interpreting the results through the lens of a simple model of genes and environment

The previous section establishes in two data sets the absence of large racial test score gaps in the first year of life, with large differences by race emerging as children age. These data by themselves, however, do not provide clear direction as to which theories most successfully explain the observed patterns in the data. To aid in that task, we construct a stylized model of the factors underlying test scores. We then calibrate the model using research-generated parameter estimates to determine which of the competing theories are most consistent with the correlations in the data

A. The model

We model test scores as determined by four factors: innate mental ability (denoted I), environment (E), and an error term composed of two parts, a person-age specific error component (θ) which is constant for an individual taking repeated versions of the test at a given age, and a random noise term (ε) which varies with each administration of the test. Both error terms are assumed to be mean zero. The person-age specific error term captures the fact that different skills are tested at different ages (at early ages, babbling is one of the measures, at later ages, how quickly one can complete a maze, e.g.). An individual may excel at the particular skills emphasized when testing is done at one age, in which case the internal validity of the test will be high, but the predictive value for future test scores (when other skills are emphasized) may be much lower. Formally, we model the test score S^* as generated by:

(1) $$S_{it}^* = \alpha_t I_{it} + \beta_t E_{it} + \varepsilon_{it} + \theta_{it},$$

where i indexes individuals and t corresponds to the age at which the test is taken. Our primary focus is on two periods t, which we denote e for "early" (age 8-12 months) and l for "later" in life (typically age 4 and beyond, but potentially as young as age two).

The test scores, measures of native ability, and environmental factors are all scaled in arbitrary units. Without loss of generality, we normalize the variables I and E to have mean zero and variance equal to one. For simplicity, we will assume in the formal modeling that there is a zero covariance between *I*, *E*, and the two error terms. We will also assume that cov(θ_{ie} ,θ_{il})=0, cov(ε_{ie} ,ε_{il})=0, and cov(θ_{it} ,ε_{it})=0 for all *t*. Further, we assume that environment and innate mental ability are uncorrelated, though we briefly discuss at the end of this section the implications for the model of allowing for feedback effects between intelligence and environment ala Dickens and Flynn (2001). We do allow for correlations between innate mental ability at early and later points in life, and likewise environment may be correlated over time for an individual. The coefficients on innate mental ability and environment are allowed to vary depending on the age at which the test is taken.

Innate mental ability and environment are not directly observable. Rather, we see only a test score. The test score is also measured in arbitrary units. We transform this test score, which already has mean zero given the assumptions above, to have variance equal to one across the population by dividing through by the standard deviation of the test, calling the resulting variable S_{it}:

(2) $$S_{it} = \frac{\alpha_t I_{it} + \beta_t E_{it} + \theta_{it} + \varepsilon_{it}}{(\alpha_t^2 + \beta_t^2 + \lambda_t^2 + \sigma_t^2)^{1/2}}$$

where λ_t^2 and σ^2 denote the variances of θ_{it} and ε_{it} respectively.

Given our assumptions above, the internal reliability of the test as measured by the correlation in test scores when the same person takes the test twice at the same age is

(3) $$corr(S_{it}, S_{it}^{'}) = \frac{\alpha_t^2 + \beta_t^2 + \lambda_t^2}{\alpha_t^2 + \beta_t^2 + \lambda_t^2 + \sigma_t^2}$$

The correlation between an individual's test scores early in life and when later tested is given by

(4) $$corr(S_{ie}, S_{il}) = \frac{\alpha_e \alpha_l \operatorname{cov}(I_e, I_l) + \beta_e \beta_l \operatorname{cov}(E_e, E_l)}{(\alpha_e^2 + \beta_e^2 + \lambda_e^2 + \sigma_e^2)^{1/2} (\alpha_l^2 + \beta_1^2 + \lambda_1^2 + \sigma_1^2)^{1/2}}.$$

Because the variance of each of the test scores is normalized to one, the correlation between the test scores is also the covariance between the test scores.

Assuming that the same equations determine Black and White test scores and that the tests are not racially biased (in the sense that θ and ε are mean zero for all races), the mean racial test score gap at a point at a given age is represented by

(5) $$\bar{S}_{W,t} - \bar{S}_{B,t} = \frac{\alpha_t (\bar{I}_{W,t} - \bar{I}_{B,t}) + \beta_t (\bar{E}_{W,t} - \bar{E}_{B,t})}{(\alpha_t^2 + \beta_t^2 + \lambda_t^2 + \sigma_t^2)^{1/2}}$$

where *W* and *B* correspond to White and Black, respectively, and a bar over a variable represents the sample mean.

B. Reconciling the empirical estimates of the parameters with the model

The model presented above, in its most general form, imposes few restrictions on the processes generating the data. For instance, an individual's innate mental ability and environment are both allowed to change over time, as is the contribution of each of those factors in determining the test score. In what follows, we consider the extent to which more restrictive versions of the model that correspond to plausible theories of nature vs. nurture are able to reproduce observed patterns in the data.

Table 4 presents the eight values that we use as stylized facts in calibrating the model.[17] The first two rows of the table correspond to the internal reliability of the test. Andreassen and Fletcher (2005) report an internal reliability of .80 for the BSID test at an early age. Later intelligence tests have been found to have internal reliability of .94 (White et al., 2004). Another parameter of interest is the correlation between test scores

[17] I cases in which the literature provides a range for particular values, we use the weighted average.

early and later in life, as captured by equation (4). Based on Figure 2, we assign a value of .30 to that correlation. Also relevant to pinning down our model is the correlation between test scores when the same individual takes the test later in life, but separated by several years. We assume a correlation of .60 in that case (Cruse et al., 1996). The estimated Black-White test score gap at young ages is taken as .06 based on our findings, compared to a gap of .80 at later ages based on our findings in CPP. The measured correlation between test scores early and late in life and parental test scores is also necessary for the analysis. Based on prior research, we take these two correlations as .30 and .39 respectively.

The primary puzzle raised by our results is the following: how does one generate small racial gaps on the BSID test scores administered at ages 8-12 months and large racial gaps in tests of mental ability later in life, despite the fact that these two test scores are reasonably highly correlated with one another (ρ=.3), and both test scores are similarly correlated with parental test scores (ρ≥.3)?

Model 1: Innate mental ability and environment are constant over time

The simplest version of the model is one in which an individual's innate mental ability and environment stay fixed over time, and the relationship between the test score and these two factors is the same both early and later in life. Mathematically, in such a model, $I_{ie} = I_{il}$, $E_{ie} = E_{il}$, $\alpha_e = \alpha_l$, and $\beta_e = \beta_l$. To solve the model, it is also convenient to assume that $\lambda_e^2 = \lambda_l^2$. Given these restrictions and noting the observed correlation in the data, equation (4) above simplifies to

$$(4') \qquad corr(S_{ie}, S_{il}) = \frac{\alpha^2 + \beta^2}{(\alpha^2 + \beta^2 + \lambda^2 + \sigma_e^2)^{1/2}(\alpha^2 + \beta^2 + \lambda^2 + \sigma_l^2)^{1/2}} = .30$$

Rearranging equation (3) and imposing the observed levels of internal reliability for the tests administered early and late in life yields

(3’) $\alpha^2 + \beta^2 + \lambda^2 + \sigma_e^2 = \dfrac{\alpha^2 + \beta^2 + \lambda^2}{.80}$

(3’’) $\alpha^2 + \beta^2 + \lambda^2 + \sigma_l^2 = \dfrac{\alpha^2 + \beta^2 + \lambda^2}{.94}$

Substituting (3’) and (3’’) into the denominator of equation (4’) eliminates σ_e^2 and σ_l^2 from (4), yielding

(6) $$\frac{\alpha^2 + \beta^2}{(\alpha^2 + \beta^2 + \lambda^2)} = .345\,.$$

In order for this simple model to reconcile the high internal reliability of the tests with the much lower correlations for an individual across ages, almost two-thirds of the variance in test scores at a point in time is being driven by the person-age specific error term, rather than by innate mental ability or environment. Since the impact of innate mental ability and environment are held fixed, the only thing that varies over time is the person-age specific error term. The necessary variance in the person-age specific error term required to fit the data would appear to be far too large to be realistic. Furthermore, in this simplest of models, the racial test score gap observed early and late in life should be similar, which is not the case. Consequently, the data are inconsistent with a simple characterization in which both the level of an individual’s innate mental ability and environment remain fixed over time and so too does the importance of those factors in determining test scores.

Model 2: Innate mental ability emerges over time, environment remains fixed

An alternative model is one in which an individual's innate mental ability has not yet fully developed at age 1, but otherwise the restrictions in the simple model are maintained. In the most extreme conceptualization of this view, one would assume that I_{ie}=0 for all individuals. Such a model can generate a small racial test score gap early in life and a one-standard deviation gap later in life if $\bar{I}_{W,l} > \bar{I}_{B,l}$. This model still implies a large variance on the person-age specific shock. Most importantly, the model cannot explain how one observes a similar degree of correlation between parental test scores and their children's test scores both early and later in life. Given the high degree of heritability in intelligence, one would expect that if early test scores are not capturing innate mental ability, then they should correlate much less strongly with parental test scores than later tests.[18]

Past research finds a correlation between children's test score at age one and parental test scores of around .30. The correlation between children tested at later ages and their parents is roughly .36. Even if all of this increased correlation between parent and child with age is attributed to the late emergence of innate mental ability, this channel can explain less than one-tenth of the racial test score gap that emerges later in life.[19]

Model 3: Innate mental ability remains fixed; environment grows in importance over time

One simple parameterization of the model that can generate results that are more consistent with the observed correlations is one in which (1) innate mental ability remains

---

[18] One possible line of argument in favor of such a model would be that early in life parents exert an enormous influence on the child's environment, and that there is a strong correlation between the parent's early environment and the child's. Arguing against this conjecture is the relatively weak impact of observable dimensions of environmental factors like SES in our test score regressions in the first year of life.

[19] This estimate is likely to overstate the true effect through this channel since the internal reliability of the children's tests rise with age. Since there is a greater noise component in the early tests, one would expect the correlation with parental scores to be lower than later in life, even if the actual correlation between the true measure of IQ between parent and child remains constant with age.

fixed for an individual over time, (2) the early and late test scores are equally influenced by innate mental ability, and (3) the importance of environmental factors grows with age.[20] A mathematical characterization of these assumptions implies $I_{ie}= I_{il}$, $\alpha_e= \alpha_l$, $\beta_e=0$, and $\beta_l>0$. As in the earlier model, for convenience we will impose $\lambda_e^2 = \lambda_l^2$, although this assumption could be relaxed. Under these assumptions equation (4) becomes

$$(4'')\ corr(S_{ie}, S_{il}) = \frac{\alpha^2}{(\alpha^2 + \lambda^2 + \sigma_e^2)^{1/2}\,(\alpha^2 + \beta_1^2 + \lambda^2 + \sigma_1^2)^{1/2}} = .30 .$$

Solving for σ_e^2 and σ_l^2 using the internal validity of the tests leads to

$$(7)\qquad corr(S_{ie}, S_{il}) = \frac{\alpha^2}{1.153*(\alpha^2 + \lambda^2)^{1/2}(\alpha^2 + \beta^2 + \lambda^2)^{1/2}} = .30 .$$

The racial test score gaps early and late in life are respectively

$$(8)\qquad \overline{S}_{W,e} - \overline{S}_{B,e} = \frac{\alpha(\overline{I}_W - \overline{I}_B)}{1.118*(\alpha^2 + \lambda^2)^{1/2}} = .06\text{, and}$$

$$(9)\qquad \overline{S}_{W,l} - \overline{S}_{B,l} = \frac{\alpha\left(\overline{I}_{W,t} - \overline{I}_{B,t}\right) + \beta_l\left(\overline{E}_W - \overline{E}_B\right)}{1.031*\left(\alpha_t^2 + \beta_l^2 + \lambda_2\right)^{1/2}} = .8$$

To close the model, we make the additional assumption that all of the variation in test scores when children take the test later in life but three years apart (such test scores are correlated .60) is due to the error terms, rather than changes in innate mental ability or environment. This assumption will tend to exaggerate the magnitude of λ^2. That assumption yields a final equation

$$(10)\ corr(S_{il}, S_{il+3}) = \frac{\alpha^2 + \beta_l^2}{1.064*(\alpha^2 + \beta_l^2 + \lambda^2)} = .60 ,$$

[20] Alternatively, one can allow for the environmental impact to be similar early and late in life, with Black and White environments being similar on average early in life, but with Blacks experiencing much worse environments later in life.

where we have already solved out for the ε term in equation (10) using the internal reliability of the test later in life.

Equations (7)-(10) contain five unknowns (α, β_I, λ, and the mean racial gaps in I and E). Without loss of generality, we can restrict α=1, with β and λ now scaled relative to α, leaving four equations and four unknowns.[21] Tedious algebraic manipulation yields a solution to this set of equations:

$$\alpha^* = 1,\ \beta^* = 1.145,\ \lambda^2 = 1.144,\ \bar{I}_W - \bar{I}_B = .102, \text{ and } \bar{E}_W - \bar{E}_B = 1.282$$

This model is able to generate the observed racial test score gaps both early and late in life, and it does so with a smaller, more realistic variance on the person-age specific error term.[22] This model can also reproduce the high levels of internal reliability of the tests, fit the .30 correlation between early and later test scores, and plausibly generate similar correlations between parent and child test scores early and later in life.[23] In this model, a one-standard deviation change in environment has roughly the same impact on test scores as does a one-standard deviation change in innate mental ability.

The required mean difference in environment between Blacks and Whites to generate the observed gap of .383 at age two in the ECLS is roughly .60 standard deviations. To generate the .80 standard deviation racial gap in CPP at ages four and seven requires an average difference in environment between Blacks and Whites of 1.282

---

[21] This can be seen by multiplying both the numerator and denominator of the left-hand sides of equations (7)-(9) by $1/\alpha$.

[22] Whereas in the earlier model the person-age specific error term had to represent more than two-thirds of the overall variance in test scores, in this model, about one-third of the variance comes from this source.

[23] Early in life, all of the variation in test scores is driven by intelligence, which is highly heritable, generating a parent-child correlation. Later in life, the intelligence channel continues to operate, but there is an additional effect operating through environment. If the cross-generation correlation in environment was as strong as that in intelligence, then one would expect to observe a higher correlation between later test scores and parental test scores than is the case for early tests. In practice, however, the cross-generation correlation in environment is much weaker than that for intelligence. Consequently, one could plausibly generate similar correlations between a parent's test score and a child's test score early and late in life in this model.

standard deviations. Fryer and Levitt (2004) find a .75 standard deviation difference between Blacks and Whites in socioeconomic status, a .83 standard deviation gap in the number of children's books in the home, a 1.30 standard deviation difference in female-headed households, a 1.51 standard deviation difference in whether or not one feels safe in their neighborhood, a 1.5 standard deviation difference in the percentage of kids in their school who participate in the free lunch program, and a 1.31 difference in the amount of loitering reported around the school by non-students.[24] Adding to this is the possible role of difficult to measure factors such as racial discrimination or behavioral differences discouraging achievement (e.g. Austen-Smith and Fryer, 2005), the required environmental gap across races of .6 to 1.28 standard deviations is extremely large, but not completely out of the realm of plausibility. Moreover, Turkheimer et al. (2003) present evidence that the impact of SES on IQ is higher at low SES levels. If that is true, than a smaller environmental gap could account for the racial test score differences since Blacks are concentrated among the lower portion of the SES distribution.

Further support for the idea that environmental factors become more important with age comes from the pattern of parameters on SES in our estimation. At age nine months in ECLS, moving from the lowest SES quintile to the highest SES quintile is associated with a test score gain of .082 standard deviations (standard error=.019). At age two, going from the lowest to highest SES quintile is associated with a .485 standard deviation jump in test scores (standard error=.032). Thus the measured impact of SES is more than five times bigger at age two than at age one in ECLS.[25] Controlling for SES

[24] All estimates are derived by taking the difference in the mean of a variable between Blacks and Whites and dividing by the standard deviation for Whites. The socioeconomic composite measure contains parental income, education, and occupation.

[25] In CPP, the measured impact of parental education also rises sharply over time. For instance, at age 8 months a child whose mother graduated from college outscores an otherwise similar child with a mother

(just one component of environment) reduces the raw Black-White test score gap at age two by nearly 40 percent.

More complicated models

In order for genetic differences across races to explain the observed patterns in the data, one needs a more complicated model than our model 2; one in which intelligence has multiple dimensions (e.g. lower and higher order thinking), only some of which differ across races and only some of which are present early in life. In this more complicated model, if only lower order thinking is relevant at age one, and there are no racial differences in lower order thinking, one can explain both absence of racial test score gaps at age one and a high observed correlation between child and parent test scores at that age. Later in life, higher order thinking emerges. If racial gaps exist in higher order thinking and parent-child correlations are similar for higher order and lower order thinking, then one can account for Blacks falling behind and child-parent correlations in test scores remaining relatively constant with age.

Another potentially important extension to the simple models presented here is to allow for positive feedback between phenotype and environment as in Dickens and Flynn (2001). This feedback introduces a multiplier effect with age which, in the presence of small initial racial differences in innate ability, lessens the magnitude of environmental differences necessary to explain the observed test score gaps later in life.

V. Discussion

---

who dropped out of high school by .081 standard deviations, whereas that same gap for a four year old child is .589 standard deviations.

The debate over racial differences in intelligence is among the most divisive in the social sciences. Utilizing a newly available, nationally representative data set with measures of mental function among children before their first birthday, we find little evidence of systematic racial differences in the first year of life. Some substantively small, but statistically significant differences are present in the raw data. Including controls for age, socio-economic status, home environment and prenatal environment largely erase these small differences. By age two, however, substantial racial gaps emerge. We document similar patterns in a second, non-nationally representative data set. A simple calibration exercise suggests that many of the basic facts in the data can be generated from a model in which there are small mean differences in intelligence across races, but large environmental differences across races that become increasingly important as children age.

Although damaging to the simplest form of the hypothesis that genetic differences are at the root of racial gaps in intelligence, the results of our analysis do not preclude a possible role for a genetic contribution to racial differences in intelligence for a number of reasons. First, one could reasonably argue that the control variables we include in the regression analysis are themselves partly genetically determined. By controlling for factors such as socio-economic status and birth weight (which systematically differ across races), we may indirectly be parsing out important channels through which genetics are operating. The fact that the raw differences in test performance across races are so small, however, makes this argument largely moot.

A second argument in defense of the genetic story is that these early tests of mental function are racially biased in favor of Blacks. This is an issue worthy of further testing, although the fact that small racial differences are observed across the wide range of skills that comprise the tests is circumstantial evidence against this claim. Furthermore, controlling for a measure of motor skills (on which Blacks outperform Whites) only slightly increases the racial gap on mental function.

A final argument in defense of the genetic story would be one in which the racial differences are concentrated in higher order thinking (or general intelligence, "g", see Jensen, 1998) which may not yet have emerged among one year olds. Unlike the arguments in the preceding paragraphs, it is not clear that such a hypothesis is easily tested – even with the passage of time.

References

Andreassen, C. and P. Fletcher. Early Childhood Longitudinal Study Birth Cohort. Psychometric Analysis. Volume 1 of the ECLS-B Methodology Report for the 9 Month Data Collection. US Department of Education. 2005. Washington, DC. National Center of Education Statistics.

Acheson, Shawn and Molfese, Victoria J. "Infant and Preschool Mental and Verbal Abilities: How Are Infant Scores Related to Preschool Scores?" *International Journal of Behavioral Development*, 1997, 20(4), pp. 595-607.

Andersson, Helle W.; Sommerfelt, Kristian and Sonnander, Karin. "Gender and its contribution to the prediction of cognitive abilities at 5 years." *Scandinavian Journal of Psychology*, 1998, 39, pp. 267–74.

Austan-Smith, David, and Roland Fryer "An Economic Analysis of 'Acting White.' *Quarterly Journal of Economics,* 2005, 120 (2), pp. 551-583.

Barnard, Kathryn E.; Bee, Helen L.; Clark, Barbara; Eyres, Sandra J.; Gray, Carol A.; Hammond, Mary A.; Snyder, Charlene and Spietz, Anita L. "Prediction of IQ and Language Skill from Perinatal Status, Child Performance, Family Characteristics and Mother-Infant Interaction." *Child Development*, 1982, 53(5), pp. 1134-56.

Baughman, Earl and Dahlstrom, Grant W. *Negro and White Children: A Psychological Study in the Rural South.* New York: Academic Press, 1968.

Bayley, Nancy. "On the Growth of Intelligence." *American Psychologist*, 1955, 10, pp.805-18.

Bethel, James; Green, James L.; Kalton, Graham and Nord, Christine. *Early Childhood Longitudinal Study, Birth Cohort (ECLS–B), Sampling*. Volume 2 of the *ECLS-B Methodology Report for the 9-Month Data Collection, 2001–02*. Washington, DC: U.S. Department of Education, NCES, 2004.

Birns, Beverly and Golden, Mark. "Prediction of intellectual performance at 3 years from infant tests and personality measures." *Merrill Palmer Quarterly*, 1972, 18, pp. 53-58.

Bracken, Bruce A.; Insko W. and Sabers, D. "Performance of Black and White Children on the Bracken Basic Concept Scale." *Psychology in Schools*, 1987, 24(1), pp. 22–27.

Bradley, Robert; Caldwell, Bettye M. and Elardo Richard. "The Relation of Infants' Home Environment to Mental Test Performance from Six to Thirty-six Months: A Longitudinal Analysis." *Child Development*, 1975, 46(1), pp. 71-76.

Brooks-Gunn, Jeanne; Duncan, Greg J.; Klebanov, Pamela K. and Sealand, Naomi. "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology*, 1993a, 99(2), pp. 353–95.

Brooks-Gunn, Jeanne; Duncan, Greg J. and Klebanov, Pamela K. "Economic Deprivation and Early-Childhood Development." *Child Development,* 1994, 65 (2), pp. 296–318.

__________. "Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment and Maternal Characteristics." *Child Development*, 1996, 67(2), pp. 396–408.

Burkett, John; McMillen, Marilyn; Owings, Jeffery and Pinkerton, Daniel B. "Statistics in Brief: Making the Cut: Who Meets Highly Selective College Entrance Criteria?" Washington, DC: National Education Longitudinal Study

(NELS),1995.
Butler, Bruce V.; Goffeney, Barbara and Henderson, Norman B. "Negro-White, Male-Female Eight-month Developmental Scores Compared with Seven-Year WISC and Bender Test Scores." *Child Development*, 1971, 42, pp. 595-604.
Campbell, Frances A.; Nicholson, Janet E. and Ramey, Craig T. "The Predictive Power of the Bayley Scale of Infant Development and the Stanford-Binet Intelligence Test in a Relatively Constant Environment." *Child Development*, 1973, 44, pp. 790-95.
Campbell, Ernest Q.; Coleman, James S.; Hobson, Carol J.; McPartland, James; Mood, Alexander M.; Weinfeld, Frederic D. and York, Robert L. *Equality of Educational Opportunity.* Washington D.C.: U.S. Government Printing Office (GOP), 1966.
Chan Toy, Colleen; Deitz, Jean; Engel, Joyce M. and Wendel, Susan. "Performance of 6-Month-Old Asian American Infants on the Movement Assessment of Infants: A Descriptive Study." *Physical & Occupational Therapy in Pediatrics: A Quarterly Journal of Developmental Therapy*, 19(3/4), pp.5-23, 2000.
Coley, Richard J. *An Uneven Start: Indicators of Inequality in School Readiness. Policy Information Report. Educational Testing Service.* Princeton, NJ: Educational Testing Service (ETS), 2002.
Colombo, John. *Infant Cognition: Predicting Later Intellectual Functioning*. Newbury Park, CA: Sage Publications Inc., 1993.
Cruse, Carol L.; Dumont, Ron; Price, Linda and Whelley, Peter. "The Relationship between the Differential Ability Scales (DAS) and the Wechsler Intelligence Scale for Children--Third Edition (WISC-III) for Students with Learning Disabilities." *Psychology in the Schools,* 1996, 33(3), pp. 203-09.
John C. DeFries, Peter McGuffin, Gerald E. McClearn, Robert Plomin. *Behavioral Genetics*. Worth Publishers; 4th edition
Dezoete, Anne J.; MacArthur, B.A. and Tuck, B. "Prediction of Bayley and Stanford-Binet Scores with a group of very low birthweight children." *Child Care, Health & Development*, 2003, 29(5), pp. 367-72.
Dickens, William T.; Flynn, James. "Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved." *Psychological Review,* 2001, 108(2): 346-369.
DiLalla, Lisabeth F.; Cypher, Lisa H.; Fagan, Joseph F.; Fulker, David W.; **Haith, Marshall M.; Thompson, L.A.; Plomin, Robert and Phillips, Kay.** "Infant Predictors of Preschool and Adult IQ: A Study of Infant Twins and Their Parents." *Developmental Psychology,* 1990, 26(5), pp. 759-69.
DiLalla, Lisabeth F.; Lovelace, Laneel and Molfese, Victoria J. "Perinatal, Home Environment and Infant Measures as Successful Predictors of Preschool cognitive and Verbal Abilities." *International Journal of Behavioral Development,* 1996, 19(1), pp. 101-19.
Enright, Mary K.; Jaskir, John and Lewis, Michael. "The Development of Mental Abilities in Infancy." *Intelligence,* 1986, 10, pp. 331-54.
Escalona, Sibylle K. and Moriarty, Alice. "Prediction of school age intelligence from infant tests." *Child Development*, 1961, 32(3), pp. 597-605.
Feldman, Judith F.; McCarton, Cecelia; Rose, Susan A. and Wallace, Ina.F.

“Information Processing at 1 Year: Relation to Birth Status and Developmental Outcome During the First 5 Years.” *Developmental Psychology,* 1991, 27(5), pp. 723-37.

Fryer, Roland and Levitt, Steven. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Review of Economics and Statistics*, 2004 86, pp. 447-64.

__________. “The Black-White Test Score Gap Through Third Grade.” *American Law and Economic Review* (forthcoming).

Fryer, Roland, and Steven Levitt. “Testing for Racial Differences in the Mental Ability of Young Children.” *NBER Working Paper No. 12066.*

Gannon, D.R. “Relationships between 8-Mo. Performance on the Bayley Scale of Infant Development and 48-Mo. Intelligence and Concept Formation Scores.” *Psychological Reports*, 1968, 23, pp. 1199-205.

Gravem, Howard; Ireton, Harold and Thwing, Edward. “Infant Mental Development and Neurological Status, Family Socioeconomic Status and Intelligence at Age Four.” *Child development,* 1970, 41(4), pp. 937-45.

Hamilton, Barton H. and McManus, Brian P. “Competition, Insurance, and Quality in the Market for Advanced Infertility Treatment.” Mimeograph. St. Louis, MO: Washington University, John M. Olin School of Business, 2004.

Hernstein, Richard J. and Murray, Charles. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press, 1994.

Honzik, Marjorie P. “Measuring mental abilities in infancy: The value and limitations,” in M. Lewis, ed., *Origins of Intelligence*. 2nd ed. New York, NY: Plenum, 1983, pp. 67-105.

Humphreys, Lloyd G. “Trends in Levels of Academic Achievement of Blacks and Other Minorities.” *Intelligence*, 1988, 12(3), pp. 231–60.

Jensen, Arthur R. *Educability and Group Differences*. New York: Free Press, 1973.

__________. *The G Factor: The Science of Mental Ability*. Westport, CT: Praeger, 1998.

Karlberg, P.; Klackenberg, G.; Engström, L.; Klackenberg-Larsson, L.; Lichtenstein, H.; Stensson, J. and Svennberg, I. “The development of children in a Swedish urbancommunity: A prospective, longitudinal study: parts l-VI.” *Acta Paediatrica Scandinavica Supplement*, 1968, 187.

Kaufman, Alan S., and Kaufman, Nadeen L. *K-ABC: Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Services, 1983.

Kopp, Claire B. and McCall, Robert B. “Predicting later mental performance for normal, at-risk, and handicapped infants,” in Paul B. Baltes and Orville G. Brim, Jr., eds., *Life-span development and behavior*. Vol 4. New York: Academic Press, 1990, pp. 31-61.

Krohn, Emily J. and Lamp, Robert E. “Current Validity of the Stanford-Binet Fourth Edition and K-ABC for Head Start Children.” *Journal of Psychology*, 1989, 27(1), pp. 59–67.

Malina, Robert M. “Racial/ethnic variation in the motor development and performance of American children.” *Canadian Journal of Sport Sciences*, 13(2), pp.136-43, 1988.

McGrath, Ellen; Bellinger, David C.; **Newburger, Jane W.; Wypij, David and Rappaport, Leonard A.** “Prediction of IQ and achievement at age 8 years from neurodevelopmental status at age 1 year in children with D-transposition of the great arteries.” *Pediatrics*, 2004, 114 (5), pp. 572-76.

McGurk, Frank C. J.; Osborne, Robert T.; Osborne, Travis R.; and Shuey, Audrey M., eds. *The Testing of Negro Intelligence*. Vol. 2. Athens, GA: The Foundation for Human Understanding, 1982.

Myrianthopoulos, Ntinos C.; Naylor, Alfred F. and Willerman, Lee. "Intellectual Development of Children from Interracial Matings: Performance in Infancy and at 4 Years." *Behavior Genetics,* 1974, 4(1), pp. 83-90.

Naglieri, Jack A. "WISC-R and K-ABC Comparison for Matched Samples of Black and White Children." *Journal of Social Psychology,* 1986, 24(1), pp. 81–88.

Neisser, Ulric; Boodoo, Gwyneth; Bouchard, Thomas J.; Boykin, A. Wade; Brody, Nathan; Ceci, Stephen J.; Halpern, Diane F.; Loehlin, John C.; Perloff, Robert, Sternberg, Robert J. et al. "Intelligence: Knowns and unknowns." *American Psychologist*, 1996, 51, pp. 77-101.

Nisbett, Richard E. "Race, Gentics and IQ," in Christopher Jencks and Meredith Phillips, eds., *The Black-White test score gap.* Washington, DC: The Brookings Institute, 1998, pp. 86-102.

Niswander, KR and M Gordon. *The women and their pregnancies: the Collaborative Perinatal Study of the National Institute of Neurological Diseases and Stroke.* Washington, DC. US Govt Print Off, 1972.

Nord, Christine; Andreassen, Carol; Branden, Laura; Dulaney, Rick; Edwards, Brad; Elmore, Anne; Flanagan, Kristin Denton; Fletcher, Philip; Green, Jim; Hilpert, Richard et al. *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), User's Manual for the ECLS-B Nine-Month Public- Use Data File and Electronic Code Book.* Washington, DC: U.S. Department of Education, NCES, 2004.

Nord, C., Edwards, B., Andreassen, C., Green, J. L., and Wallner-Allen, K. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), User's Manual for the ECLS-B Longitudinal 9-Month–2-Year Data File and Electronic Codebook (NCES 2006–046). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2006.

Phillips, Meredith; Brooks-Gunn, Jeanne; Crane, Jonathan; Duncan, Greg J. and Klebanov, Pamela K. "Family Background, Parenting Practices, and the Black-White Test Score Gap," in Christopher Jencks and Meredith Phillips, eds., *The Black-White test score gap.* Washington, DC: The Brookings Institute, 1998, pp. 103-45

Phillips, Meredith. "Understanding Ethnic Differences in Academic Achievement: Empirical Lessons from National Data," in David W. Grissmer and Michael J. Ross, eds., *Analytic Issues in the Assessment of Student Achievement.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES), 2000, pp 103-32.

Rose, Susan A. and Wallace, Ina F. "Visual recognition memory: a predictor of later cognitive functioning in preterms." *Child Development,* 1985, 56 (4), pp. 843-52.

Rushton, Philippe J. "Race and crime: international data for 1989-1990." *Psychological Reports*, 1995, 76(1), pp. 307-12.

Rushton, J. Philippe and Arthur Jensen. "Thirty Years of Research on Race Differences in Cognitive Ability." *Psychology, Public Policy, and Law,* 2005, 11(2), pp. 235-94.

Scarr, Sandra. *Race, Social Class and Individual Differences in I.Q.* Mahwah, NJ: Lawrence Erlbaum Associates, 1981.

Scott, R. and Sinclair, D. "Ethnic-related cognitive profiles of black and white pre-school children." *Homo- Journal of Comparative Human Biology*, 1997, 28, pp.116-20.

Shuey, Audrey M. *The Testing of Negro Intelligence.* Lynchburg, VA: J.P. Bell Company, 1958.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D'Onofrio, and Irving I. Gottesman. "Socioeconomic status modifies heritability of iq in young children." Psychological Science, 2003, 14 (6), pp.623-628.

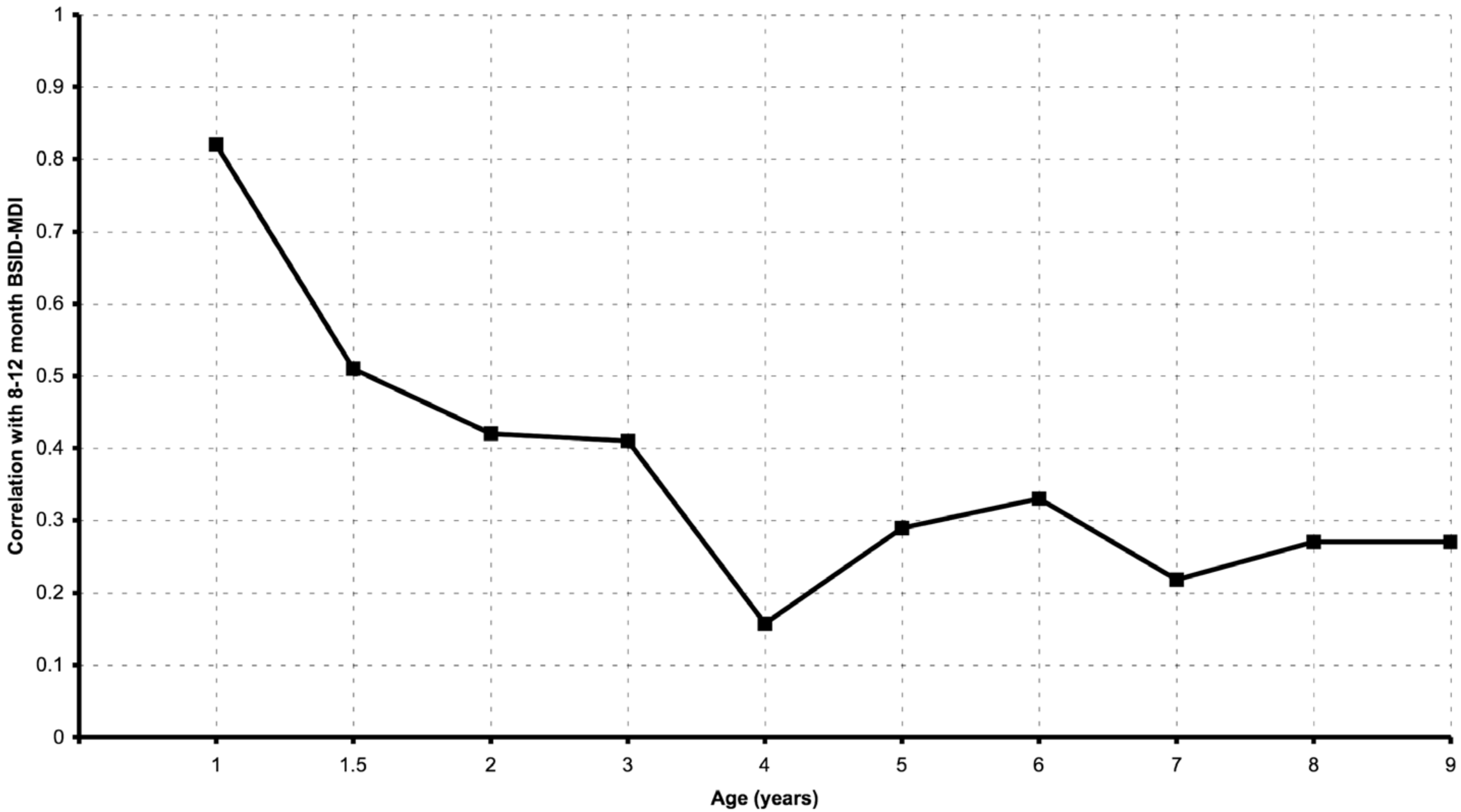
White, Roberta, Richard Cambell, Patricia Janulewicz. *Neuropsychological Assessments in Children from a Longitudinal Perspective for the National Children's Study*. Battelle Memorial Institute, 2004.

Wilson, Ronald S. "The Louisville Twin Study: Developmental Synchronies in Behavior." *Child Development,* 1983, 54(2), pp. 293-316.

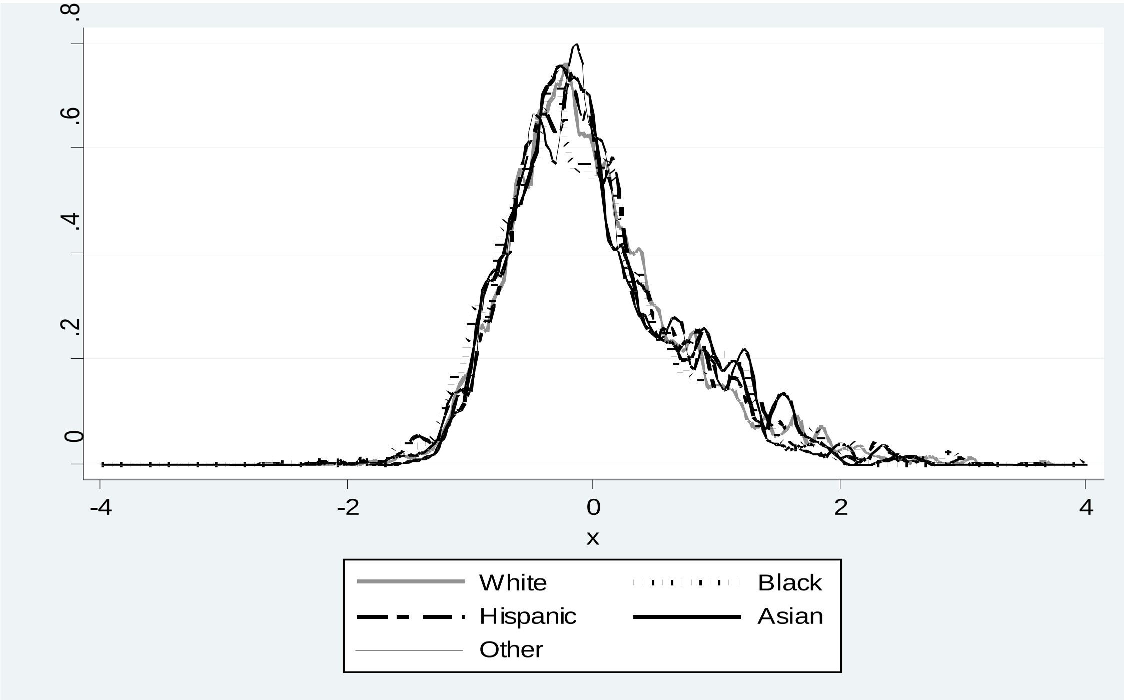
Wilson, Ronald S. and Matheny, Adam P. "Mental Development: Family Environment and Genetic Influences." Intelligence, 1983, 7, pp. 195-215.

Yeates, K. O.; MacPhee, D.; Campbell, F. A. and Ramey, C. T. "Maternal IQ and home environment as determinants of early childhood intellectual competence: Developmental analysis." *Developmental Psychology*, 1983, 19, pp. 731-9.

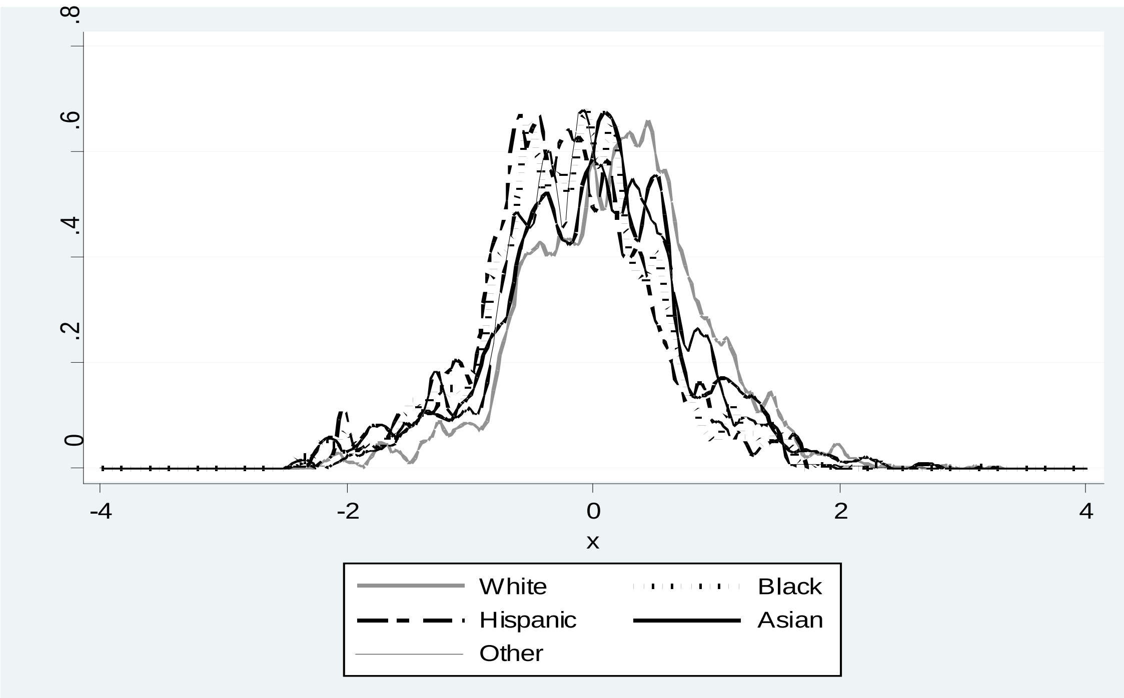
Figure 1



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Correlation with 8-12 month BSID-MDI
1
1.5
2
3
4
5
6
7
8
9
Age (years)

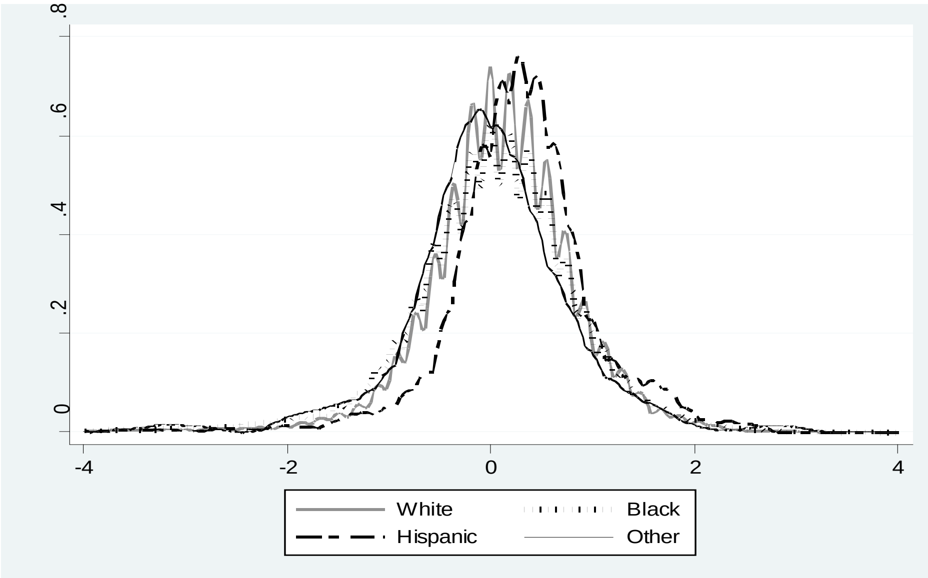


(A)

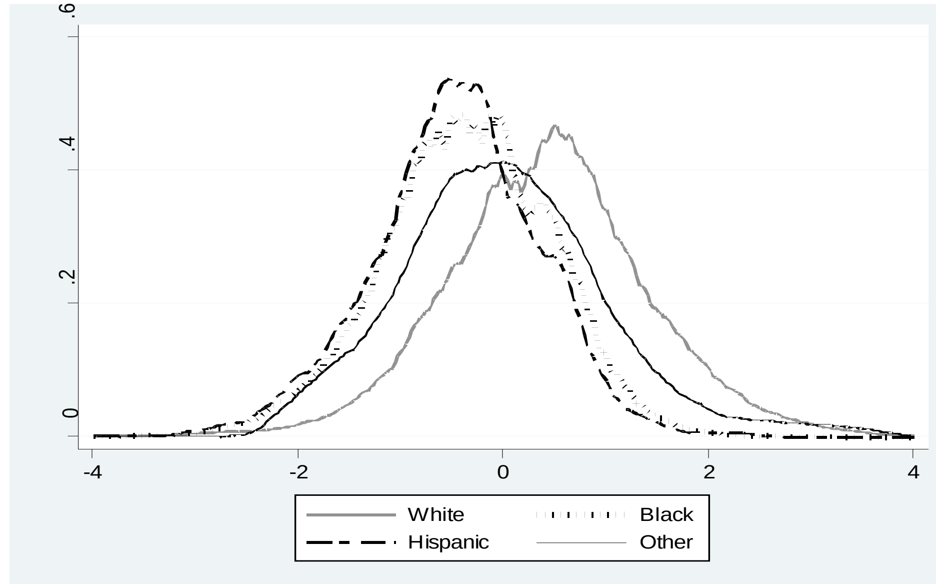


(B)

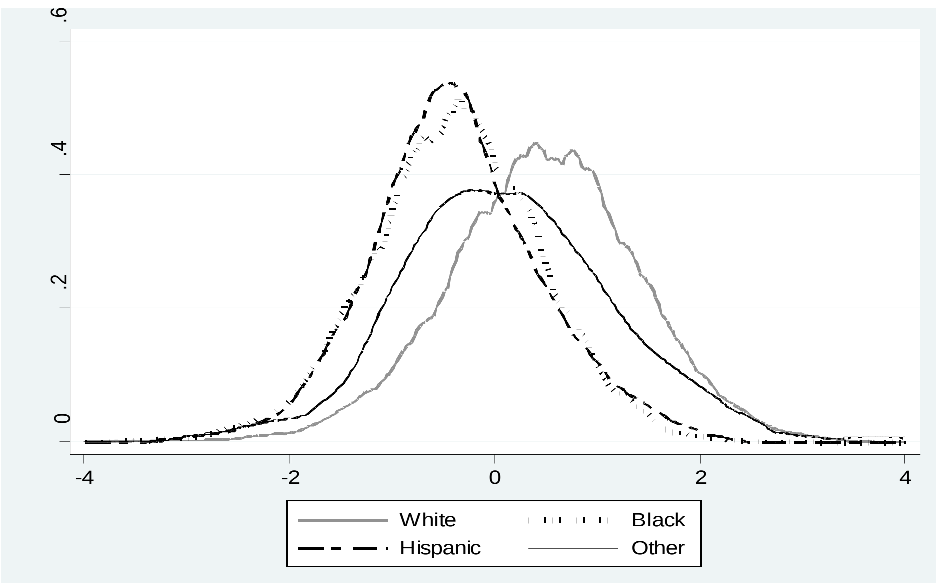
Figure 2:
Distribution of Test Scores in the ECLS for 9 month-olds (A) and 2 year-olds (B).



(A)



(B)



(C)

Figure 3:
Distribution of Test Scores in the CPP for 9 month-olds (A), 4 year-olds, (B) and 7 year-olds (C).

Table 1a: Summary statistics from ECLS-B Data Set

Variables	All Races	White	Black	Hispanic	Asian	Other
Cognitive Development						
Mental function composite score in first wave (age 8-16 months)	0.000	0.013	-0.042	-0.005	0.006	-0.009
	(0.998)	(0.938)	(0.927)	(0.909)	(1.051)	(1.112)
Mental function composite score in second wave (age 23-26 months)	0.000	0.176	-0.207	-0.246	-0.039	-0.059
	(0.996)	(0.908)	(0.902)	(0.856)	(1.006)	(1.153)
Demographics						
age in first wave (in months)	10.002	9.977	10.007	10.039	10.086	10.025
	(2.605)	(2.437)	(2.358)	(2.449)	(2.337)	(2.775)
age in second wave (in months)	23.868	23.846	23.868	23.916	23.878	23.849
	(1.232)	(1.103)	(1.165)	(1.240)	(1.097)	(1.514)
female	0.489	0.489	0.479	0.493	0.475	0.501
region:						
northeast	0.169	0.185	0.152	0.144	0.196	0.137
midwest	0.224	0.283	0.198	0.107	0.161	0.261
region:	0.364	0.351	0.570	0.315	0.200	0.275
west	0.244	0.180	0.080	0.433	0.442	0.327
Home environment						
SES quintiles:						
1st quintile (lowest)	0.188	0.078	0.337	0.361	0.088	0.184
2nd quintile	0.193	0.150	0.229	0.264	0.115	0.262
3rd quintile	0.201	0.208	0.213	0.188	0.133	0.200
4th quintile	0.208	0.267	0.137	0.122	0.155	0.207
5th quintile (highest)	0.210	0.297	0.085	0.065	0.509	0.147
number of siblings	0.991	0.962	1.106	1.013	0.772	1.009
	(1.455)	(1.279)	(1.455)	(1.461)	(1.312)	(1.910)
Family Structure:						
biological mother and biological father present	0.799	0.888	0.413	0.802	0.938	0.733
one biological parent present	0.185	0.092	0.580	0.183	0.059	0.247
one biological parent and one non-biological parent present	0.012	0.015	0.005	0.009	0.001	0.012
other parental configuration	0.005	0.005	0.002	0.006	0.001	0.007
mother's age	29.970	28.596	25.139	26.182	29.850	26.055
	(17.197)	(7.521)	(7.670)	(7.335)	(7.252)	(10.456)
parent as teacher score	29.970	31.450	28.851	27.522	27.643	30.075
	(17.197)	(15.240)	(15.838)	(16.967)	(18.212)	(22.574)
Prenatal environment						
birthweight:						
< 1500	0.012	0.010	0.024	0.011	0.007	0.014
> 1500 & < 2500	0.062	0.055	0.102	0.055	0.062	0.058
> 2500 & < 3500	0.537	0.507	0.593	0.557	0.697	0.549
3500 or more	0.389	0.428	0.280	0.377	0.234	0.378
Percent premature	0.114	0.099	0.173	0.113	0.082	0.149
	(0.363)	(0.312)	(0.422)	(0.341)	(0.301)	(0.573)
Days premature (conditional on being premature)	21.042	20.828	22.610	20.485	16.328	21.270
	(23.494)	(23.857)	(23.792)	(19.831)	(13.506)	(33.527)
multiple birth:						
singleton	0.966	0.963	0.968	0.976	0.983	0.944
twin	0.030	0.035	0.032	0.022	0.017	0.027

higher order	0.002	0.002	0.001	0.001	0.000	0.000
Number of observations	8,806	3,793	1,394	1,780	888	951

NOTES: Table entries are sample means (with standard deviations in parentheses for those variables that take on a value other than zero or one) of child-level data for those children in ECLS-B who do not have missing values for test scores in waves one and two. Test scores are a mental composite score, normalized to have a mean of 0 and standard deviation of 1. The category "White" includes only non-Hispanic Whites. Precise definitions of the variables are provided in the data appendix. The SES measure incorporates information on parental education, occupational status, and family income. Days premature is conditional on being premature. In all cases, sample weights provided with ECLS are used in the calculations. All variables other than second wave test score and age at second wave are values at the time of the first wave. Values are missing for some variables for some children; these missing values are not included in the calculation of the sample means in this table. In the regression analysis, indicators for missing values are included as controls.

Table 1b: Summary statistics for CPP Data Set

Variables	All Races	White	Black	Hispanic	Other
Cognitive Development					
Mental function composite score in first wave (age 8 months)	0.000	0.044	-0.051	0.227	-0.127
	(1.000)	(0.925)	(1.060)	(0.925)	(1.292)
Mental function composite score in second wave (age 45-55 months)	0.000	0.424	-0.361	-0.471	-0.019
	(1.000)	(1.000)	(0.843)	(0.814)	(1.033)
Mental function composite score in third wave (age 81-99 months)	0.000	0.457	-0.397	-0.389	0.112
	(1.000)	(0.957)	(0.857)	(0.840)	(1.092)
Demographics					
Age in first wave:					
< 7.5 months	0.006	0.004	0.008	0.006	0.004
7.5 - 8.5 months	0.835	0.811	0.856	0.856	0.807
8.5 - 9 months	0.087	0.105	0.070	0.081	0.090
9 - 10 months	0.066	0.074	0.059	0.053	0.076
> 10 months	0.006	0.006	0.006	0.004	0.022
age in second wave (in months)	48.216	48.389	48.035	48.497	48.655
	(1.422)	(1.302)	(1.491)	(1.578)	(1.462)
age in third wave (in months)	84.314	84.335	84.286	84.446	84.489
	(2.493)	(2.042)	(2.802)	(3.256)	(2.209)
female	0.501	0.495	0.508	0.489	0.507
Home environment					
Father's education:					
dad is high school drop out	0.502	0.396	0.613	0.781	0.500
dad has high school degree	0.314	0.308	0.324	0.188	0.224
dad has some college	0.090	0.127	0.053	0.016	0.031
dad has at least a college degree	0.094	0.169	0.011	0.016	0.245
Father's occupation:					
dad has no occupation	0.004	0.001	0.007	0.000	0.000
dad has a professional occupation	0.171	0.269	0.065	0.036	0.245
dad has a non-professional occupation	0.825	0.730	0.928	0.964	0.755
Mother's education:					
mom is high school drop out	0.555	0.423	0.678	0.918	0.625
mom has high school degree	0.323	0.371	0.282	0.055	0.175
mom has some college	0.072	0.111	0.035	0.014	0.075
mom has at least a college degree	0.050	0.095	0.005	0.014	0.125
Mother's occupation:					
mom has no occupation	0.133	0.057	0.205	0.301	0.183
mom has a professional occupation	0.074	0.130	0.021	0.000	0.108
mom has a non-professional occupation	0.792	0.813	0.774	0.699	0.708
Income in first three months of pregnancy ($1963):					
income less than 499	0.157	0.081	0.230	0.186	0.196
income between 500 and 999	0.379	0.262	0.492	0.600	0.473
income between 1000 and 1499	0.238	0.292	0.185	0.157	0.196
income between 1500 and 1999	0.120	0.183	0.059	0.029	0.107
income between 2000 and 2499	0.057	0.094	0.022	0.014	0.018
income more than 2500	0.049	0.088	0.012	0.014	0.009
number of siblings	2.812	2.422	3.182	2.863	2.724
	(2.231)	(1.783)	(2.526)	(2.742)	(2.069)

biological mother and biological father present	0.729	0.851	0.583	0.875	0.639
mother's age	30.049	30.836	29.287	27.625	31.203
	(7.033)	(6.469)	(7.510)	(5.414)	(6.328)
mother's reaction towards child:					
indifferent	0.256	0.254	0.265	0.120	0.293
accepting	0.997	0.997	0.998	0.999	1.000
attentive	0.190	0.205	0.182	0.106	0.209
over caring	0.032	0.040	0.026	0.015	0.065
other	0.002	0.062	0.001	0.000	0.005
Prenatal environment					
birthweight:					
< 1500	0.007	0.005	0.010	0.001	0.000
> 1500 & < 2500	0.099	0.070	0.127	0.087	0.078
> 2500 & < 3500	0.657	0.617	0.693	0.700	0.620
3500 or more	0.237	0.308	0.170	0.212	0.302
Percent premature	0.819	0.843	0.795	0.855	0.841
	(0.385)	(0.364)	(0.403)	(0.353)	(0.367)
Weeks premature (conditional on being premature)	2.745	2.796	2.696	2.747	-2.756
	(0.795)	(0.771)	(0.819)	(0.701)	(0.698)
multiple birth:					
singleton	0.982	0.983	0.981	0.988	0.969
twin	0.017	0.017	0.018	0.012	0.018
higher order	0.001	0.000	0.000	0.000	0.013
Number of observations	31,116	14,335	15,667	891	223

NOTES: The entries are means and standard deviations of child-level data for those children in CPP who do not have missing values for the given variable. Only children with test scores in all three waves of the study are included. Test scores are a mental composite score, normalized to have a mean of 0 and standard deviation of 1. The category "White" includes only non-Hispanic Whites. Precise definitions of the variables are provided in the data appendix. Days premature is conditional on being premature. Income is for the first three months of pregnancy in 1963 dollars. All variables other than second and third wave test scores and ages are values at the time of the first wave. Values are missing for some variables for some children; these missing values are not included in the calculation of the sample means in this table. In the regression analysis, indicators for missing values are included as controls.

Table 2: Estimating Group Differences in the Mental Function Composite Score in the First Wave of the ECLS-B

	(Children approximately 9 months in age)					
	Dependent variable: Standardized Mental Function Composite Score					
	(1)	(2)	(3)	(4)	(5)	(6)
black	-0.055	-0.072	-0.062	-0.035	-0.014	0.015
	[0.029]	[0.032]*	[0.017]**	[0.018]*	[0.019]	[0.018]
hispanic	-0.018	0.001	-0.037	-0.005	-0.004	-0.007
	[0.026]	[0.031]	[0.016]*	[0.017]	[0.017]	[0.017]
asian	-0.007	0.026	-0.017	-0.016	-0.013	-0.01
	[0.038]	[0.042]	[0.023]	[0.023]	[0.023]	[0.022]
other	-0.022	-0.021	-0.032	-0.014	-0.008	0.000
	[0.039]	[0.040]	[0.024]	[0.024]	[0.024]	[0.023]
controls:						
Interviewer fixed effects	No	Yes	Yes	Yes	Yes	Yes
age, gender, region	No	No	Yes	Yes	Yes	Yes
SES	No	No	No	Yes	Yes	Yes
home environment	No	No	No	No	Yes	Yes
prenatal conditions	No	No	No	No	No	Yes
Observations	8,806	8,806	8,806	8,806	8,806	8,806
R-squared	0.00	0.14	0.71	0.71	0.71	0.76

Robust standard errors in brackets

* significant at 5%; ** significant at 1%

NOTES: The dependent variable is the mental composite score, which are normalized to have a mean of 0 and a standard deviation of 1 in the full, unweighted sample. Non-Hispanic Whites are the omitted race category, so all of the race coefficients are gaps relative to that group. The unit of observation is a child. robust standard errors are in parentheses. Estimation is done using weighted least squares, using sample weights provided in the data set. As one moves from left to right in the table, additional controls are added to the specification. In addition to the variables included in the table, indicator variables for children with missing values on each covariate are also included in the regressions. * significant at 5%; ** significant at 1%

Table 3: Estimating Group Differences in the Mental Function Composite Score

	Across waves of the ECLS and CPP Data Sets									
	Dependent variable: Standardized Mental Function Composite Score									
	ECLS				CPP					
	9 months		2 years		8 months		4 years		7 years	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
black	-0.055	0.015	-0.383	-0.213	-0.096	0.024	-0.785	-0.296	-0.854	-
	[0.029]	[0.018]	[0.028]**	[0.032]**	[0.012]**	[0.017]	[0.011]**	[0.016]**	[0.010]**	[0
hispanic	-0.018	-0.007	-0.422	-0.249	0.183	-0.039	-0.895	-0.542	-0.846	-
	[0.026]	[0.017]	[0.025]**	[0.029]**	[0.034]**	[0.040]	[0.032]**	[0.039]**	[0.031]**	[0
asian	-0.007	-0.01	-0.215	-0.294	-	-	-	-	-	
	[0.038]	[0.022]	[0.037]**	[0.036]**	-	-	-	-	-	
other	-0.022	0.000	-0.235	-0.132	-0.171	-0.107	-0.443	-0.271	-0.345	-
	[0.039]	[0.023]	[0.040]**	[0.038]**	[0.067]*	[0.060]*	[0.062]**	[0.057]**	[0.061]**	[0
controls:										
Interviewer fixed effects	No	Yes	No	Yes	No	Yes	No	Yes	No	
age, gender, region	No	Yes	No	Yes	No	Yes	No	Yes	No	
SES	No	Yes	No	Yes	No	Yes	No	Yes	No	
home environment	No	Yes	No	Yes	No	Yes	No	Yes	No	
prenatal conditions	No	Yes	No	Yes	No	Yes	No	Yes	No	
Observations	8,806	8,806	8,806	8,806	31,116	31,116	31,116	31,116	31,116	3
R-squared	0.00	0.75	0.07	0.31	0.00	0.24	0.15	0.33	0.18	

Robust standard errors in brackets

* significant at 5%; ** significant at 1%

NOTES: The dependent variable is the mental composite score, which is normalized to have a mean of 0 and a standard deviation of 1 in each wave of the full, unweig sample. Non-Hispanic Whites ire the omitted race category, so are all of the race coefficients are relative to that group. The unit of observation is a child. Standard erro parentheses. Estimation is done using weighted least squares for the ECLS in columns (1)-(4) using sample weights provided in the data set. Estimation is done using leasat squares for the CPP sample in columns (5)-(10).   In addition to the variables included in the table, indicator variables for children with missing values on each co are also included in the regressions.

Table 4: Estimates Used in Calibrating the Model

Measure	Algebraic Representation	Value Used	Source
Internal Reliability of BSID	$corr(S_{ie}, S'_{ie})$	0.08	Andreassen and Fletcher (2005)
Internal Reliability of later test scores	$corr(S_{il}, S'_{il})$	0.94	White et. al. 2004
Correlation between BSID and later test scores	$corr(S_{ie}, S'_{il})$	0.3	Figure 2
Correlation between later IQ tests when taken three years apart	$corr(S_{il}, S'_{il+3})$	0.6	Cruse et.al. 1996
Black-White test score gap on BSID	$\overline{S}_{W,e} - \overline{S}_{B,e}$	0.06	Table 2
Black-White test score gap on later tests	$\overline{S}_{W,l} - \overline{S}_{B,l}$	0.8	Table 3
Correlation between parental IQ and BSID	$corr(S_{ie}, S'_{iP})$	0.36	Yeates et al. 1983
Correlation between parental IQ and child IQ on later tests	$corr(S_{il}, S'_{P})$	0.39	Yeates et al. 1983

Appendix Table 1: Literature on the Relationship Between Bayley Scores and Future IQ

Author(s)	Age of Bayley (mos)	Future IQ Test and Age (years)	Correlation	N
Gannon, 1968	8	Stanford-Binet at age 4	0.13	371
Butler, Goffeney & Henderson, 1971	8	Wechsler Intelligence Scale for Children at age 7	0.19	626
Ireton, Gravem & Thwing, 1970	8	Stanford-Binet at age 4	0.25	500
Myrianthopoulos, Naylor & Willerman, 1974	8	Stanford-Binet at age 4	0.22	129
Bayley, 1949	7 to 9	Bayley at 10, 11 & 12 months	0.81	40
Bayley, 1955	9 & 12	Mean 16-18 months IQ	0.32, 0.30	45
Ramey et al., 1973	9 to 12	Stanford-Binet at age 3	0.71	11
Wilson, 1983	9 & 12	Bayley at ages 1.5 & 2 years	0.51, 0.42	340
		Stanford-Binet at age 3	0.38	340
		Wechsler Preschool and Primary Scale of Intelligence at ages 4, 5, & (	0.29, 0.35, 0.33	340
		Wechsler Intelligence Scale for Children at ages 7, 8, & 9	0.27, 0.27, 0.27	340
Bradley, Caldwell & Elardo, 1975	12	Stanford-Binet at age 3	0.32	77
Barnard, et al.1982	12	Stanford-Binet at age 4	0.21	156
Enright, Jaskir & Lewis, 1986	12	Stanford-Binet at age 3	0.12	116
DiLalla et al., 1990	12	Stanford-Binet at age 3	0.32	40
Rose et al., 1991	12	Stanford-Binet at ages 3 & 4	0.30, 0.22	40
DiLalla, Lovelace & Molfese, 1996	12	Stanford-Binet at ages 3 & 4	0.17, 0.14	94
Acheson & Molfese, 1997	12	Stanford-Binet at ages 3, 4 & 5	0.14, 0.15, 0.06	89

NOTES: To generate relevant literature, Pubmed, EBSCOhost and JSTOR were searched for all years. Keywords used for the search included, “Bayley Scale of Infant Development,” “BSID, “predictive power,” “correlation,” and “IQ” – which generated a list of roughly 2,000 references. Among these, we selected according to the following criteria. (1) the test of infant cognition was the Bayley Scale of Infant Development (BSID. The correlation at 10 to 12 months is from the California First-Year Mental Scale, the first version of the BSID-MDI); (2) The “future” test age was administered more than 2 years after the infant measure was taken (with the exception of the 'future ages' of 1, 1.5 and 2 years); (3) the initial BSID had to be administered between 6 and 12 months of age; (4) the sample had to be “representative” – we omitted studies with subjects afflicted by major health issues (e.g., heart disease, Down syndrome, mental retardation) or severe complications at birth (i.e. pre-term, mother used drugs or alcohol); and (5) only studies done in the United States were used. The final compilation contained 14 studies.

Appendix Table 2: Estimating Group Differences in the Mental Function Composite Score for 9 month olds

	Dependent variable: Standardized Mental Function Composite Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.055	-0.072	-0.062	-0.035	-0.014	0.015
	[0.029]	[0.032]*	[0.017]**	[0.018]*	[0.019]	[0.018]
Hispanic	-0.018	0.001	-0.037	-0.005	-0.004	-0.007
	[0.026]	[0.031]	[0.016]*	[0.017]	[0.017]	[0.017]
Asian	-0.007	0.026	-0.017	-0.016	-0.013	-0.01
	[0.038]	[0.042]	[0.023]	[0.023]	[0.023]	[0.022]
Other	-0.022	-0.021	-0.032	-0.014	-0.008	0
	[0.039]	[0.040]	[0.024]	[0.024]	[0.024]	[0.023]
Age						
9			0.331	0.329	0.325	0.332
			[0.014]**	[0.014]**	[0.014]**	[0.014]**
10			0.668	0.667	0.665	0.694
			[0.016]**	[0.016]**	[0.016]**	[0.016]**
11			0.966	0.966	0.959	1.008
			[0.023]**	[0.023]**	[0.023]**	[0.022]**
12			1.357	1.355	1.349	1.387
			[0.027]**	[0.027]**	[0.026]**	[0.025]**
13			1.646	1.646	1.636	1.662
			[0.027]**	[0.027]**	[0.027]**	[0.026]**
14			1.839	1.841	1.823	1.843
			[0.051]**	[0.051]**	[0.049]**	[0.049]**
15			2.118	2.12	2.106	2.145
			[0.048]**	[0.048]**	[0.047]**	[0.044]**
16+			2.45	2.45	2.442	2.457
			[0.078]**	[0.079]**	[0.078]**	[0.077]**
female			0.045	0.045	0.046	0.056
			[0.011]**	[0.011]**	[0.011]**	[0.011]**
Socioeconomic Status Quintiles:						
2				0.01	0.002	0.008
				[0.018]	[0.017]	[0.017]
3				0.063	0.038	0.036
				[0.018]**	[0.018]*	[0.018]*
4				0.07	0.037	0.027
				[0.019]**	[0.020]	[0.020]
5 (highest)				0.082	0.041	0.032
				[0.019]**	[0.022]	[0.021]
Number of Siblings:						
1					-0.058	-0.057
					[0.013]**	[0.013]**
2					-0.075	-0.062
					[0.016]**	[0.016]**
3					-0.139	-0.111
					[0.024]**	[0.024]**
4					-0.152	-0.128
					[0.041]**	[0.040]**
5					-0.174	-0.111
					[0.054]**	[0.054]*

6	-0.083	-0.093
	[0.054]	[0.051]
Family configuration:		
Single biological parent	-0.033	-0.029
	[0.017]	[0.016]
Biological parent and other parent	-0.07	-0.043
	[0.057]	[0.055]
Other parent type	-0.168	-0.092
	[0.078]*	[0.085]
Midwest	-0.012	-0.004
	[0.048]	[0.047]
South	0.069	0.079
	[0.050]	[0.049]
West	0.085	0.083
	[0.065]	[0.065]
Mother's age	-1.082	-1.186
	[0.609]	[0.581]*
Mother's age^2	0.076	0.081
	[0.043]	[0.041]*
Mother's age^3 (*100000)	-260.187	-269.603
	[149.246]	[142.322]
Mother's age^4 (*100000)	4.286	4.344
	[2.510]	[2.393]
Mother's age^5 (*100000)	-0.027	-0.027
	[0.016]	[0.016]
Mother's age missing	-5.83	-6.426
	[3.342]	[3.178]*
Parentscore	0.167	0.101
	[1.194]	[1.181]
Parentscore^2	-0.01	-0.005
	[0.078]	[0.078]
Parentscore^3 (*100000)	34.783	13.898
	[251.323]	[250.012]
Parentscore^4 (*100000)	-0.612	-0.241
	[3.947]	[3.936]
Parentscore^5 (*100000)	0.004	0.002
	[0.024]	[0.024]
Missing Parentscore	1.321	1.019
	[7.089]	[6.986]
Birthweight:		
>= 1500 & < 25000		0.371
		[0.039]**
>= 2500 & < 3500		0.592
		[0.040]**
>= 3500		0.642
		[0.040]**
Days Premature:		
0		0.136
		[0.073]
7		0.074
		[0.078]
14		-0.046

						[0.080]
21						0.007
						[0.083]
28						-0.197
						[0.090]*
35						-0.042
						[0.090]
42						-0.174
						[0.115]
49						-0.191
						[0.094]*
56						-0.017
						[0.098]
63						-0.181
						[0.090]*
70						-0.466
						[0.107]**
77						-0.509
						[0.095]**
Singleton birth						0.091
						[0.094]
Twin birth						0.007
						[0.094]
Triplet or more birth						0
						[0.000]
Constant	0.013	0.01	-0.654	-0.711	3.876	4.143
	[0.015]	[0.015]	[0.014]**	[0.020]**	[7.892]	[7.692]
Observations	8806	8806	8806	8806	8806	8806
R-squared	0	0.14	0.71	0.71	0.71	0.76

Robust standard errors in brackets

* significant at 5%; ** significant at 1%

omitted categories are northeast, birthweight missing, days premature missing, missing info on type of birth, no siblings, two biological parents

Note that the triplet or more variable is dropped from the last specification because it is perfectly collinear with the Mother's Age Missing variable (it's collinear with the type of birth variables and needs to be dropped and the mother's age missing, so something needs to be dropped)

NOTES: See notes to Table 2. The columns in this table report the full regression results of specifications (1) through (6) in Table 2.