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

Roland G. Fryer, Jr.
Harvard University Society of Fellows
and NBER

and

Steven D. Levitt
University of Chicago and
American Bar Foundation

February 2006

* We would like to thank Lawrence Katz, Kevin Murphy, Andrei Shleifer, and Lawrence H. Summers for helpful discussions on this topic. Katherine Barghaus, Sheldon Bond, Walter Lamberson and Ethan Lieber 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 Fryer, Littauer Building, 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.

Abstract

On tests of intelligence, Blacks systematically score worse than Whites, whereas Asians frequently outperform 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. The only statistically significant racial difference is that Asian children score slightly worse than those of other races. To the extent that there are any genetically-driven racial differences in intelligence, these gaps must either emerge after the age of one, or operate along dimensions not captured by this early test of mental cognition.

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 typically 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 one standard deviation racial gap observed later in life is present by age three (Jensen and Rushton, 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 .7, see, for instance, Neisser, 1996) and persistent racial gaps in test scores is evidence of genetic differences across races (Jensen, 1973, 1998; Jensen and Rushton, 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 that the inclusion of a small number of demographic and socio-economic controls erases the gap (Fryer and Levitt, 2004). Although through the first four years of school, the Black-White test score gap grows substantially, 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 9-12 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 genetics story cannot be definitively rejected with these data.

Because of data limitations, prior research has not been able to directly address this question (Jensen and Rushton 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 (McGarth 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 the assumption that these tests are strongly related to 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 controls only for the child's age and gender, the mean White infant outscores the mean Black infant by .064 standard deviation units – only a tiny fraction of the one standard deviation racial gap observed at older ages. The raw scores for Blacks are indistinguishable from Hispanics and Asians, who also slightly under perform Whites. Adding controls for 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, except Asians. In our sample Asian babies do slightly, but statistically significantly, worse than babies of other races.

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.

II. The Data

The 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 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 eight to eleven months of age in five 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, early problem solving (e.g., when a toy is out of reach, using another object as a tool to retrieve the toy), and communicating with words. 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 no direct evidence regarding the correlation between performance on this exact 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 observed in this prior literature, weighting the estimates of the various papers by the number of subjects included in each study.[4] The horizontal axis is the age in years that the

[3] These children will eventually be resurveyed four times between the present and first grade. At the present time, only the first wave of data is available.
[4] Appendix Table 1 list the full set of papers and reported correlations on which Figure 1 is based.

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. 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).[5]

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. Campbell et al. (1983), for instance, report 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 as .36. 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 blacks, a correlation of .36 between child's BSID and mother's IQ would imply expected mean differences between white and black one-year olds of .36 standard deviations.[6]

The ECLS-B sample includes observations on 10,688 children. For 556 of these individuals, no mental ability test was performed. These subjects are dropped from the

[5] 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).

[6] Indeed, given the results cited in Jensen and Rushton (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 one standard deviation.

analysis. This is the only exclusion we make from the sample.[7] Throughout the analysis, the results we report are weighted to be nationally representative using sampling weights included in the data set.[8]

Table 1 provides summary statistics by major racial/ethnic group in ECLS-B. 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 .018 standard deviations better than the sample mean on the overall mental measure.

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

[7] 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.

[8] 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.

observation of parent-child interactions in a structured problem-solving environment.[9] 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 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).

Figure 2 plots the density of BSID test scores by race. The test score distributions for infants of different races are essentially indistinguishable visually. Note that this is in stark contrast to the distribution of other measures of intelligence for older children in prior studies. Figure 2 makes the main point of the analysis. In the regression analysis that follows, we control for a wide range of other factors, but the basic conclusion suggested by Figure 2 is unchanged.

---

[9] 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).

III. Racial Differences in the Mental Ability of Young Children

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

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

where i indexes individuals and r corresponds to the racial group to which an individual belongs. *MENTAL TEST* reflects either the overall test score, or one of the underlying components of the test – depending on our particular specification. 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.[10]

The basic results for the coefficients on the race variables 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 includes only interviewer-fixed effects. As in the raw data, Blacks, Hispanics, Asians, and "Other" races slightly under perform Whites. Only for Blacks can one reject the null of no difference at the .05 level.

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. Column (2) adds controls for age at which the test is administered and the gender of the child. Because the age the test is

[10] 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).

taken is such an important determinant of test performance, we adopt an extremely flexible, non-parametric functional form for age (including age in days linearly, age in days squared, all the way up to a quintic in age). The R-squared in the regression jumps substantially with the inclusion of the age and gender controls (from .14 in column 1 to .68 in column 2); which is mainly due to the age controls. A child taking the test at age eleven months is predicted to perform .66 standard deviations better than the same child when tested at age nine months, according to our estimates. Girls outscore boys by .06 standard deviations on average, a difference that persists throughout all our specifications. Including these controls improves the performance of Whites versus other races because Whites are slightly younger on average when tested.

Column (3) 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 region dummies) shrinks the gap between Whites and each of the other races. The only race coefficient that remains statistically significant is the underperformance by Asians of roughly .06 standard deviations. Adding a control for the interviewer assessment of "parent as teacher" in column (5) further shrinks all of the race coefficients.

The final column adds controls for a range of prenatal condition variables (birth weight, premature birth, and multiple births). The coefficient on Black becomes very close to zero; the other race variables are not greatly affected. All the point estimates are substantively small: the largest gap, between Asians and Whites, is only .048 standard deviation units. After Whites, the group performing best is Blacks, although one cannot reject equality between Blacks and any of the other groups.

IV. Sensitivity Analysis and Extensions of the Basic Analysis

Although on average we observe only minor differences across races in mental test performance, important differences in sub-groups of the population remain a possibility. Table 3 presents a wide range of estimates to test the sensitivity of our basic results. Each row of the table presents the race coefficients from a separate regression. In each case, the specification uses the full set of controls included in the final column of the preceding table.

The top row of the table simply reproduces the baseline estimates from column (6) of Table 2. The next row presents unweighted estimates. The subsequent rows segment the table by gender, SES quintile, family structure, region, urbanicity, and birth weight. Of the 104 coefficients presented, twelve are statistically significant at the .05 level; almost all of these correspond to underperformance by Asians.

Evidence exists that there are racial differences in early motor skills, with Blacks outperforming other races and Asians lagging (Malina, 1988; Colombo, 1993; Toy et al. 2000). In our data, both Blacks and Asians outperform Whites on the test of motor skills. If the early life test of mental abilities depends on both sensorimotor and intellectual skills, precocity in the former might mask a deficiency in the latter among Blacks. Therefore, we report results controlling for an individual's performance on the test of motor skills**.** The black coefficient increases from -.002 to -.034 with the inclusion of motor skills, though the absolute magnitude of the difference continues to be extremely small. All other racial groups remain essentially unchanged.

The final sensitivity test we report corresponds to the age at which the test of mental function is administered. In prior studies, the correlation between BSID tests and later IQ scores were higher when the BSID was administered to older infants (Bayley, 1955*)*. Thus, if true, racial differences in intelligence are being masked by limitations of the BSID to ascertain these differences for children at young ages, then we would expect the racial differences in our sample to be smaller among the youngest children sampled and larger among the older children sampled. The final rows of Table 3 provides no evidence of this for Blacks, although for the "other race" category, the gaps are steadily rising versus Whites with age.

Investigating the individual components underlying the composite mental test score

The overall test score is made up of five underlying components: exploration, exploration with purpose, babbling, problem solving, and word use. Table 4 presents regression estimates on each of these parts of the test. Each column corresponds to a different component of the test. The mean level of proficiency and standard deviation in that mean are reported at the bottom of each column. Note that we have not transformed the proficiency scores on the sub-tests to have mean zero and standard deviation one, so the estimates in this table are not directly comparable to those of the two preceding tables. Because there is so little variation on some of the sub-tests, transforming each of the tests to have the same standard deviation across children will artificially inflate test score differences on the sections with little variation (i.e., explores objects and uses words). Only the race coefficients are shown in the tables. The specifications include the full set of controls, mirroring column (6) of Table 2. Nineteen of the twenty race coefficients are negative, implying that in virtually all of the cases the omitted group (Whites) shows greater proficiency. The magnitudes of the estimates, however, are quite small: less than one-half a percentage point difference in proficiency in almost all cases. In only one instance (less babbling done by Asians infants) is the gap between another group and Whites statistically significant at the .05 level. The only positive estimate in the table is associated with greater (but not statistically significantly different) babbling by Blacks. Thus, there is little in the underlying test components to alter the conclusions suggested by the composite measure of mental function.

V. Discussion

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. 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.

Comparing the magnitude of differences observed in these data to prior studies of children, teens, and adults puts into perspective how small the observed differences are in this analysis of infants. Phillips et al. (1998), which investigates test outcomes among kindergarteners in the early cohorts of the Children of the National Longitudinal Survey of Youth (NLSY), finds some of the *smallest* published racial gaps. They report a raw Black-White gap of over one standard deviation, which shrinks to one-third of a standard deviation with the inclusion of myriad controls. Those gaps are an order of magnitude larger than what we find among infants.[11]

Although damaging to 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 possible argument one can make against our findings is that the particular form of the BSID used in ECLS has not been proven to correlate with future measures of intelligence. The relationship between other forms of the BSID and later test scores calls this argument into question. Nonetheless, the answer to this inquiry cannot be definitively determined in this data set at the present time. In principle, however, retesting the subjects as they age provides a means of resolving this question.

A third 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 argument in the preceding paragraph, it is not clear that such a hypothesis is easily tested even with the passage of time.

---

[11] As noted earlier, one possible explanation for the growing racial gap with age is a genetics-environment interaction as proposed by Dickens and Flynn (2001).

References

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.

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.

Campbell, Frances A.; MacPhee, David; Ramey, Craig T. **and Yeates, Keith O.** "Maternal IQ and home environment as determinants of early childhood intellectual competence: Developmental analysis." *Developmental Psychology*, 1983, 19, pp. 731-39.

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.

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).

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.

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

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.

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.

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.

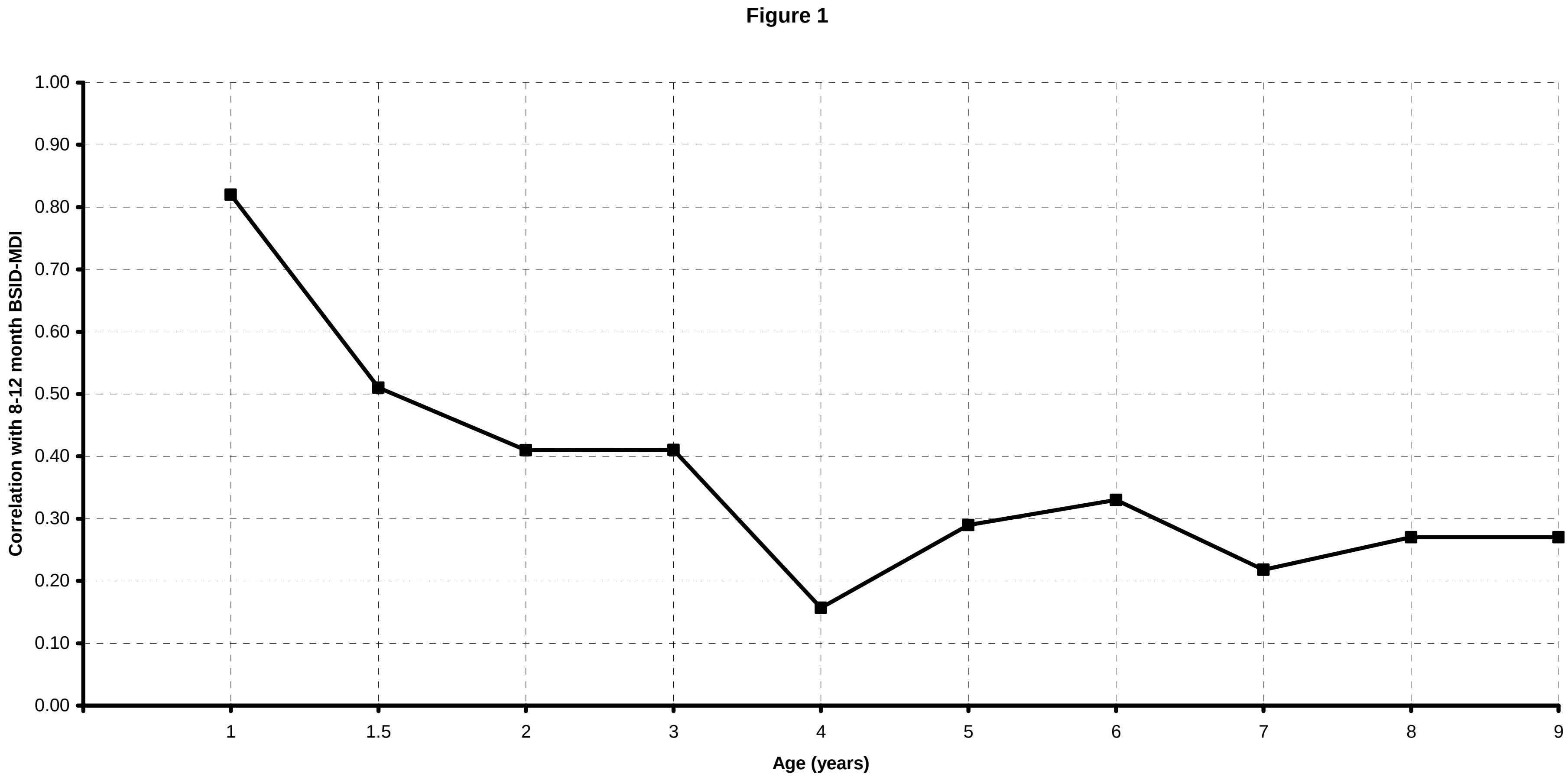
Scarr, Sandra. *Race, Social Class and Individual Differences in I.Q.* Mahwah, NJ: Lawrence Erlbaum Associates, 1981.

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.

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

Figure 1: The Correlation between Bayley Scores and Later IQ



Notes: See Appendix Table 1 for a list of sources from which the correlations are based. Each is a (sample size) weighted average of the correlations found across all relevant studies.

Figure 2: The Distribution of Bayley Scores, by Race

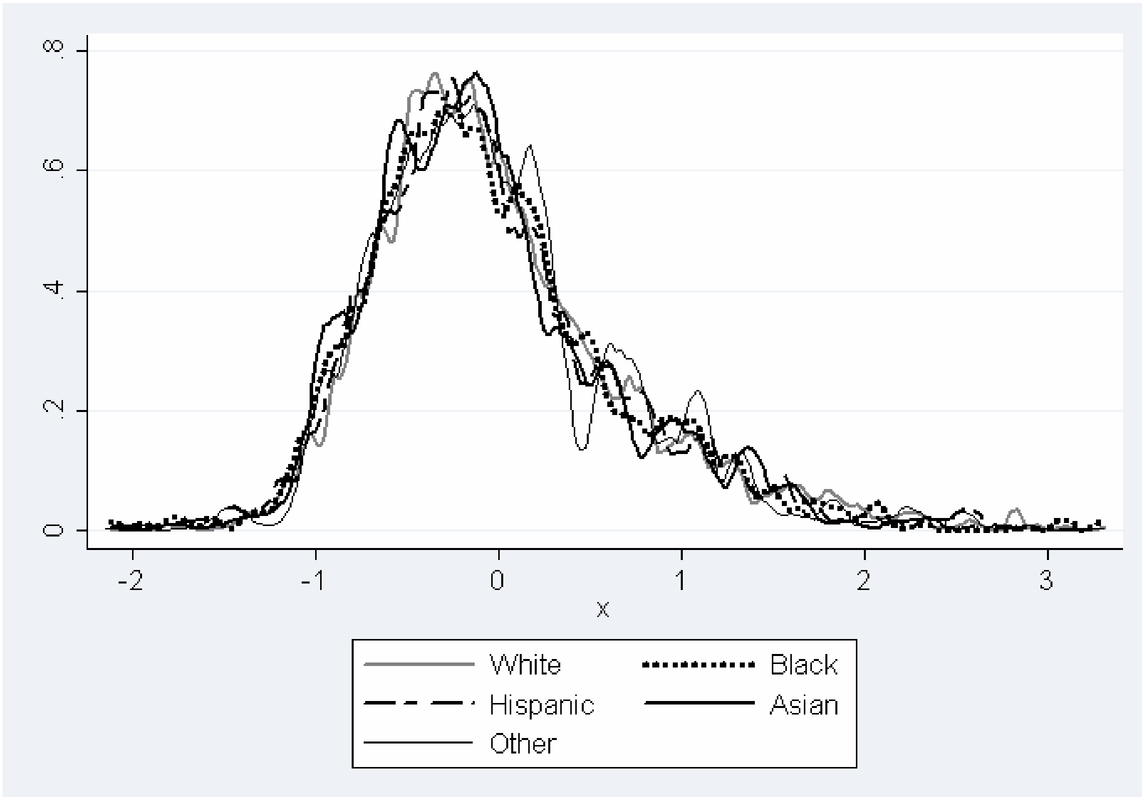


Table 1: Summary statistics

Variables	All Races	White	Black	Hispanic	Asian	Other
Cognitive Development						
Mental function composite score	0.0002	0.0182	-0.0258	-0.0185	-0.0513	0.0065
	(0.9996)	(0.9426)	(0.8951)	(0.9156)	(0.9551)	(1.1528)
Demographics						
age	10.2677	10.2379	10.2920	10.2983	10.3397	10.3285
	(1.9362)	(1.7914)	(1.7777)	(1.822)	(1.8556)	(2.4305)
female	0.4892	0.4848	0.4855	0.4987	0.4708	0.5130
	(0.6835)	(0.6401)	(0.6232)	(0.627)	(0.6593)	(0.8661)
region:						
northeast	0.1684	0.1818	0.1575	0.1470	0.1974	0.1428
	(0.5513)	(0.5355)	(0.4856)	(0.4609)	(0.6072)	(0.7097)
midwest	0.2223	0.2818	0.2092	0.1076	0.1583	0.2527
	(0.5511)	(0.558)	(0.4997)	(0.3649)	(0.4284)	(0.7413)
south	0.3675	0.3601	0.5569	0.3147	0.1876	0.2906
	(0.6566)	(0.6132)	(0.6237)	(0.5854)	(0.5275)	(0.8)
west	0.2418	0.1763	0.0765	0.4307	0.4567	0.3140
	(0.5779)	(0.4788)	(0.3479)	(0.6199)	(0.651)	(0.7526)
Home environment						
SES quintiles:						
1st quintile (lowest)	0.2000	0.0850	0.3424	0.3789	0.1021	0.1753
	(0.5346)	(0.3471)	(0.5894)	(0.6086)	(0.4213)	(0.6472)
2nd quintile	0.1999	0.1572	0.2408	0.2633	0.1417	0.2622
	(0.5391)	(0.4543)	(0.5325)	(0.5529)	(0.5695)	(0.7867)
3rd quintile	0.1999	0.2087	0.2091	0.1817	0.1436	0.2097
	(0.5456)	(0.5174)	(0.511)	(0.482)	(0.4305)	(0.6977)
4th quintile	0.2005	0.2624	0.1289	0.1146	0.1538	0.2047
	(0.5589)	(0.569)	(0.4212)	(0.3985)	(0.4298)	(0.6962)
5th quintile (highest)	0.1997	0.2868	0.0788	0.0615	0.4588	0.1480
	(0.5554)	(0.5849)	(0.339)	(0.3013)	(0.6531)	(0.5923)
number of siblings	0.9848	0.9568	1.1316	0.9899	0.7783	0.9754
	(1.4761)	(1.2918)	(1.543)	(1.46)	(1.2742)	(1.8726)
biological mother and biological father present	0.7892	0.8833	0.4111	0.7901	0.9212	0.7322
	(0.5352)	(0.4051)	(0.6171)	(0.5085)	(0.4085)	(0.7824)
one biological parent present	0.1938	0.0960	0.5793	0.1950	0.0754	0.2504
	(0.5155)	(0.3716)	(0.6185)	(0.4939)	(0.4058)	(0.7737)
one biological parent and one non-biological parent present	0.0114	0.0142	0.0067	0.0094	0.0015	0.0106
	(0.1455)	(0.1508)	(0.1007)	(0.1109)	(0.0349)	(0.1715)
other parental configuration	0.0056	0.0065	0.0030	0.0054	0.0020	0.0069
	(0.1042)	(0.0979)	(0.0578)	(0.1097)	(0.0424)	(0.056)
mother's age	27.2587	28.3941	25.1252	26.0201	29.4158	25.8535
	(8.3082)	(7.6037)	(7.5755)	(7.334)	(7.2416)	(10.6263)
parent as teacher score	29.6979	31.2523	28.4809	27.3848	27.0006	29.9106
	(17.4207)	(15.4744)	(16.1934)	(16.909)	(19.1218)	(23.2064)

Prenatal environment						
birthweight:						
< 1500	0.0127	0.0102	0.0240	0.0121	0.0072	0.0154
	(0.0445)	(0.0348)	(0.0635)	(0.0419)	(0.054)	(0.0749)
> 1500 & < 2500	0.0621	0.0557	0.0998	0.0550	0.0651	0.0601
	(0.1934)	(0.1616)	(0.2524)	(0.1773)	(0.2618)	(0.2609)
> 2500 & < 3500	0.5451	0.5123	0.6023	0.5669	0.6832	0.5458
	(0.6776)	(0.6391)	(0.5985)	(0.6153)	(0.6079)	(0.8588)
3500 or more	0.3796	0.4212	0.2732	0.3656	0.2445	0.3778
	(0.6737)	(0.6367)	(0.5794)	(0.608)	(0.5815)	(0.8539)
Percent premature	0.1167	0.1016	0.1685	0.1169	0.0948	0.1534
	(0.3677)	(0.3186)	(0.412)	(0.3492)	(0.3246)	(0.5675)
Days premature (conditional on being premature)	21.1239	21.0457	22.7998	20.2129	17.4876	21.4877
	(23.6229)	(24.2518)	(23.5968)	(19.9711)	(13.4793)	(32.8074)
multiple birth:						
singleton	0.9668	0.9624	0.9693	0.9768	0.9786	0.9452
	(0.1007)	(0.0997)	(0.091)	(0.0749)	(0.1175)	(0.1813)
twin	0.0298	0.0347	0.0300	0.0214	0.0197	0.0252
	(0.0927)	(0.0925)	(0.0882)	(0.0718)	(0.1125)	(0.113)
higher order	0.0018	0.0028	0.0007	0.0007	0.0016	0.0005
	(0.0296)	(0.0337)	(0.0214)	(0.0102)	(0.0332)	(0.0118)

NOTES: The entries are means and standard deviations of child-level data for those children in ECLS-B who do not have missing values for test scores. Test scores are a mental composite score, normalized to have a mean of 0 and a 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. The total number of children in the sample who receive a positive weight in the estimation is 10,132. In all cases, sample weights provided with ECLS are used in the calculations.

Table 2: Estimating Group Differences in the Mental Function Composite Score

	Dependent variable: Standardized Mental Function Composite Score					
	(1)	(2)	(3)	(4)	(5)	(6)
black	-0.061	-0.063	-0.037	-0.026	-0.022	-0.002
	[0.028]*	[0.017]**	[0.018]*	[0.019]	[0.019]	[0.019]
hispanic	-0.013	-0.046	-0.015	-0.02	-0.011	-0.012
	[0.028]	[0.016]**	[0.017]	[0.017]	[0.017]	[0.017]
asian	-0.026	-0.068	-0.063	-0.072	-0.056	-0.048
	[0.035]	[0.020]**	[0.020]**	[0.020]**	[0.020]**	[0.020]*
other	-0.013	-0.039	-0.024	-0.026	-0.023	-0.016
	[0.039]	[0.025]	[0.025]	[0.025]	[0.025]	[0.024]
controls:						
age, gender, region		yes	yes	yes	yes	yes
SES			yes	yes	yes	yes
home environment				yes	yes	yes
parents' score					yes	yes
prenatal conditions						yes
Observations	10132	10132	10132	10132	10132	10132
R-squared	0.14	0.68	0.68	0.68	0.68	0.71

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. Standard errors are in parentheses. Estimation is done using weighted least squares, using sample weights provided in the data set. 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. All regressions include interviewer fixed effects.

Table 3: Sensitivity Analysis of Group Differences in Mental Function Composite Score

	black	hispanic	asian	other
baseline	-0.002	-0.012	-0.048	-0.016
	[0.019]	[0.017]	[0.020]*	[0.024]
unweighted	0.020	0.000	-0.061	0.007
	[0.014]	[0.013]	[0.015]**	[0.015]
males only	-0.018	0.007	-0.024	-0.041
	[0.028]	[0.023]	[0.030]	[0.039]
females only	0.021	-0.012	-0.073	0.004
	[0.024]	[0.025]	[0.027]**	[0.029]
SES quintile 1	-0.046	-0.042	-0.143	0.001
	[0.049]	[0.044]	[0.062]*	[0.059]
SES quintile 2	-0.045	-0.059	-0.095	0.028
	[0.035]	[0.036]	[0.055]	[0.066]
SES quintile 3	-0.006	0.000	-0.019	-0.060
	[0.042]	[0.036]	[0.047]	[0.043]
SES quintile 4	0.028	0.022	-0.108	-0.028
	[0.045]	[0.039]	[0.050]*	[0.042]
SES quintile 5	0.043	-0.008	-0.017	-0.034
	[0.051]	[0.048]	[0.035]	[0.045]
if both biological parents present	0.009	-0.020	-0.034	0.009
	[0.024]	[0.019]	[0.020]	[0.027]
if a biological parent not present	0.000	0.021	-0.168	-0.043
	[0.035]	[0.039]	[0.075]*	[0.049]
if northwest	-0.010	-0.010	0.018	-0.090
	[0.045]	[0.046]	[0.060]	[0.088]
if midwest	0.113	0.033	0.019	0.029
	[0.041]**	[0.040]	[0.048]	[0.044]
if south	-0.053	-0.037	-0.070	-0.031
	[0.026]*	[0.027]	[0.035]*	[0.042]
if west	0.063	-0.005	-0.076	-0.009
	[0.052]	[0.030]	[0.031]*	[0.036]
if urban, inside UA	0.011	0.010	-0.037	-0.013
	[0.021]	[0.019]	[0.021]	[0.026]
if urban, outside UC	-0.050	-0.072	-0.015	-0.030
	[0.082]	[0.054]	[0.076]	[0.072]
if rural	-0.034	-0.117	-0.045	-0.041
	[0.057]	[0.069]	[0.224]	[0.053]
if normal birthweight	-0.008	-0.012	-0.047	-0.023
	[0.020]	[0.018]	[0.021]*	[0.026]
if moderately low birthweight	-0.005	-0.030	-0.079	-0.043
	[0.041]	[0.040]	[0.059]	[0.063]
if very low birthweight	0.071	0.009	-0.212	-0.079
	[0.046]	[0.054]	[0.149]	[0.101]
controlling for motor skills	-0.034	-0.005	-0.066	-0.037
	[0.018]	[0.016]	[0.018]**	[0.023]
age < 9 months	-0.059	-0.046	-0.054	0.037
	[0.038]	[0.033]	[0.038]	[0.048]
9 months <= age < 10 months	-0.005	0.000	-0.019	0.002
	[0.025]	[0.024]	[0.025]	[0.032]
10 months <= age < 11 months	0.024	-0.022	-0.040	-0.002
	[0.034]	[0.032]	[0.035]	[0.047]
11 months <= age < 12 months	0.003	0.014	-0.105	-0.044
	[0.064]	[0.063]	[0.070]	[0.120]
age >= 12 months	-0.014	0.039	-0.038	-0.071
	[0.072]	[0.063]	[0.066]	[0.081]

NOTES: Specifications in this table are variations on those reported in column (6) of Table 2. Only the race coefficients are reported. The top row simply reproduces the baseline results in column (6) of Table 2. The remaining rows correspond to different weights, socio-economic quintiles and other particular subsets of the data. For further details of the baseline specification, see the notes to Table 2.

Table 4: Estimating Group Differences in the Underlying Components of the Mental Function Composite Score

	explores objects	explores objects purposefully	babbles	early problem	uses words
black	-0.0007	-0.0012	0.0022	-0.0027	-0.0033
	[0.0005]	[0.0039]	[0.0047]	[0.0046]	[0.0039]
hispanic	-0.0006	-0.0039	-0.0021	-0.0013	-0.0013
	[0.0004]	[0.0039]	[0.0043]	[0.0041]	[0.0033]
asian	-0.0004	-0.0054	-0.0128	-0.0088	-0.0048
	[0.0004]	[0.0044]	[0.0053]*	[0.0047]	[0.0030]
other	-0.0002	-0.0015	-0.0031	-0.0041	-0.0031
	[0.0005]	[0.0051]	[0.0064]	[0.0061]	[0.0046]
controls:					
age, gender, region, mother's age	yes	yes	yes	yes	yes
SES	yes	yes	yes	yes	yes
home environment	yes	yes	yes	yes	yes
parents' score	yes	yes	yes	yes	yes
prenatal conditions	yes	yes	yes	yes	yes
mean of dependent variable	0.9938	0.9058	0.5471	0.0906	0.0317
sd of dependent variable	(0.0145)	(0.1555)	(0.268)	(0.2172)	(0.1399)
Observations	10132	10132	10132	10132	10132
R-squared	0.21	0.43	0.73	0.64	0.44

NOTES: Entries are unadjusted mean scores on specific components of the mental composite score. They are proficient probability scores, which are constructed using IRT scores and provide the probability of mastery of a specific skill set.

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

Author(s)	Age of Bayley (months)	Future IQ Test and Age (years)	Correlation	N
Gannon, 1968	8	Stanford-Binet at age 4	0.13	371
Butler, Goffeney and Henderson and, 1971	8	Wechsler Intelligence Scale for Children at age 7	0.19	626
Ireton, Gravem and Thwing, 1970	8	Stanford-Binet at age 4	0.25	500
Myrianthopoulos, Naylor and Willerman, 1974	8	Stanford-Binet at age 4	0.22	129
Bayley, 1955	9 & 12	Mean 16-18 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	Stanford-Binet at age 3	0.38	340
		Wechsler Preschool and Primary Scale of Intelligence at ages 4, 5, and 6	0.29, 0.35, 0.33	340
		Wechsler Intelligence Scale for Children at ages 7, 8, and 9	0.27, 0.27, 0.27	340
Bradley, Caldwell and 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 and 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 and 4	0.30, 0.22	40
DiLalla, Lovelace and Molfese, 1996	12	Stanford-Binet at ages 3 and 4	.17, .14	94
Acheson and Molfese, 1997	12	Stanford-Binet at ages 3, 4 and 5	.14, .15, .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); (2) The "future" test age was administered more than 2 years after the infant measure was taken (correlations at 12 and 18 months were obtained from Wilson 1983); (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

	Dependent variable: Standardized Mental Function Composite Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.061	-0.063	-0.037	-0.026	-0.022	-0.002
	[0.028]*	[0.017]**	[0.018]*	[0.019]	[0.019]	[0.019]
Hispanic	-0.013	-0.046	-0.015	-0.02	-0.011	-0.012
	[0.028]	[0.016]**	[0.017]	[0.017]	[0.017]	[0.017]
Asian	-0.026	-0.068	-0.063	-0.072	-0.056	-0.048
	[0.035]	[0.020]**	[0.020]**	[0.020]**	[0.020]**	[0.020]*
Other	-0.013	-0.039	-0.024	-0.026	-0.023	-0.016
	[0.039]	[0.025]	[0.025]	[0.025]	[0.025]	[0.024]
Age		488.089	483.958	498.736	505.248	504.52
		[110.641]**	[110.111]**	[109.414]**	[108.913]**	[107.444]**
Age^2		-96.648	-95.941	-98.855	-100.032	-100.028
		[21.419]**	[21.318]**	[21.180]**	[21.086]**	[20.798]**
Age^3		9.525	9.465	9.751	9.856	9.867
		[2.062]**	[2.053]**	[2.039]**	[2.031]**	[2.002]**
Age^4		-0.467	-0.464	-0.478	-0.483	-0.484
		[0.099]**	[0.098]**	[0.098]**	[0.097]**	[0.096]**
Age^5		0.009	0.009	0.009	0.009	0.009
		[0.002]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
female		0.061	0.061	0.063	0.061	0.069
		[0.011]**	[0.011]**	[0.011]**	[0.011]**	[0.011]**
Socioeconomic Status Quintiles:						
2			0.024	0.021	0.018	0.021
			[0.018]	[0.018]	[0.018]	[0.018]
3			0.076	0.067	0.055	0.052
			[0.018]**	[0.019]**	[0.019]**	[0.019]**
4			0.081	0.073	0.054	0.047
			[0.019]**	[0.021]**	[0.021]**	[0.021]*
5 (highest)			0.079	0.074	0.048	0.042
			[0.020]**	[0.022]**	[0.023]*	[0.022]
Number of Siblings:						
1				-0.066	-0.064	-0.063
				[0.013]**	[0.013]**	[0.013]**
2				-0.079	-0.079	-0.07
				[0.017]**	[0.016]**	[0.017]**
3				-0.135	-0.134	-0.117
				[0.025]**	[0.025]**	[0.025]**
4				-0.14	-0.13	-0.114
				[0.044]**	[0.044]**	[0.044]**
5				-0.102	-0.081	-0.042
				[0.059]	[0.059]	[0.057]
6				-0.122	-0.126	-0.136
				[0.077]	[0.075]	[0.073]
Family configuration:						
Single biological parent				-0.019	-0.017	-0.013
				[0.017]	[0.017]	[0.017]
Biological parent and other parent				-0.039	-0.051	-0.039
				[0.053]	[0.053]	[0.052]

Other parent type	-0.044	-0.051	0.002
	[0.110]	[0.108]	[0.111]
Midwest	-0.038	-0.035	-0.031
	[0.048]	[0.047]	[0.047]
South	0.043	0.058	0.06
	[0.056]	[0.056]	[0.055]
West	0.016	0.02	0.011
	[0.063]	[0.063]	[0.063]
Mother's age	-0.14	-0.118	-0.07
	[0.669]	[0.666]	[0.660]
Mother's age^2	0.012	0.01	0.005
	[0.047]	[0.047]	[0.046]
Mother's age^3 (*100000)	-43.527	-35.908	-13.402
	[160.510]	[159.846]	[158.698]
Mother's age^4 (*100000)	0.744	0.598	0.146
	[2.674]	[2.665]	[2.648]
Mother's age^5 (*100000)	-0.005	-0.004	0
	[0.017]	[0.017]	[0.017]
Mother's age missing	-0.506	-0.41	-0.22
	[3.718]	[3.697]	[3.664]
Parentscore		-0.269	-0.3
		[1.226]	[1.220]
Parentscore^2		0.015	0.017
		[0.080]	[0.080]
Parentscore^3 (*100000)		-38.583	-48.659
		[256.303]	[256.023]
Parentscore^4 (*100000)		0.485	0.669
		[4.012]	[4.014]
Parentscore^5 (*100000)		-0.002	-0.004
		[0.025]	[0.025]
Mparentscore		-1.741	-1.897
		[7.321]	[7.253]
Birthweight:			
< 1500			-0.355
			[0.090]**
>= 1500 & < 25000			-0.051
			[0.087]
>= 2500 & < 3500			0.068
			[0.086]
>= 3500			0.132
			[0.086]
Days Premature:			
0			0.103
			[0.060]
7			0.032
			[0.066]
14			-0.022
			[0.069]

21						0.075
						[0.070]
28						-0.097
						[0.078]
35						0.005
						[0.076]
42						0.005
						[0.087]
49						-0.061
						[0.076]
56						0.045
						[0.085]
63						-0.059
						[0.075]
70						-0.291
						[0.083]**
77						-0.365
						[0.074]**
Singleton birth						-0.001
						[0.060]
Twin birth						-0.077
						[0.060]
Triplet or more birth						-
						-
Constant	0.013	-982.802	-973.364	-1,002.51	-1,015.15	-1,012.18
	[0.014]	[227.402]**	[226.276]**	[224.953]**	[224.352]**	[221.337]**
Observations	10132	10132	10132	10132	10132	10132
R-squared	0.14	0.68	0.68	0.68	0.68	0.71

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