The Review *of* Economics *and* Statistics

VOL. LXXXVI MAY 2004 NUMBER 2

UNDERSTANDING THE BLACK-WHITE TEST SCORE GAP IN THE FIRST TWO YEARS OF SCHOOL

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

Abstract—In previous research, a substantial gap in test scores between white and black students persists, even after controlling for a wide range of observable characteristics. Using a newly available data set (the Early Childhood Longitudinal Study), we demonstrate that in stark contrast to earlier studies, the black-white test score gap among incoming kindergartners disappears when we control for a small number of covariates. Real gains by black children in recent cohorts appear to play an important role in explaining the differences between our findings and earlier research. The availability of better covariates also contributes. Over the first two years of school, however, blacks lose substantial ground relative to other races. There is suggestive evidence that differences in school quality may be an important part of the explanation. None of the other hypotheses we test to explain why blacks are losing ground receive any empirical backing.

I. Introduction

THE black-white test score gap is a robust empirical regularity. A simple comparison of mean test scores typically finds black students scoring roughly 1 standard deviation below white students on standardized tests. Even after controlling for a wide range of covariates including family structure, socioeconomic status, measures of school quality, and neighborhood characteristics, a substantial racial gap in test scores persists.[1]

Gaining a better understanding of the underlying causes of the test score gap is of great importance. Neal and Johnson (1996) and O'Neill (1990) find that most of the observed black-white wage differentials among adults disappears once lower eighth grade test scores among blacks are taken into account. Thus, eliminating the test score gap that arises by the end of junior high school may be a critical component of reducing racial wage inequality.[2]

A wide variety of possible explanations for the test-score gap have been put forth. These explanations include differences in genetic make-up (Hernstein & Murray, 1984; Jensen, 1973, 1998), differences in family structure and poverty (Armor, 1992; Brooks-Gunn & Duncan, 1997; Mayer, 1997; Phillips et al., 1998a), differences in school quality (Cook & Evans, 200), racial bias in testing or teachers' perceptions (Delpit, 1995; Ferguson, 1998; Rodgers & Spriggs, 1996), and differences in culture, socialization, or behavior (Cook & Ludwig, 1998; Fordham & Ogbu, 1986; Fryer, 2002; Steele & Aronson, 1998). The appropriate public policy choice (if any) to address the test score gap depends critically on the underlying source of the gap.

In this paper, we utilize the Early Childhood Longitudinal Study kindergarten cohort (ECLS-K) to shed new light on the test score gap. ECLS-K is a new data set administered by the Department of Education. The survey covers a sample of more than 20,000 children entering kindergarten in the fall of 1998. An enormous amount of information is gathered for each individual, including family background, school, and neighborhood characteristics, teacher and parent assessments, and test scores. The original sample of students has subsequently been reinterviewed in the spring of kindergarten and first grade.

The results we obtain using these new data are informative and in some cases quite surprising. As in previous data sets, we observe substantial racial differences in test scores in the raw data: black kindergartners score on average 0.64 standard deviation worse than whites. In stark contrast to earlier studies (including those looking at kindergartners), however, after controlling for a small number of other observable characteristics (children's age, child's birth weight, a socioeconomic status measure, WIC participation, mother's age at first birth, and number of children's books in the home), we essentially eliminate the black-white test score gap in math and reading for students entering kindergarten.[3] Controlling for a much larger set of characteristics yields the same conclusion. This

Received for publication July 5, 2002. Revision accepted for publication June 30, 2003.

* Harvard Society of Fellows and NBER; and University of Chicago and American Bar Foundation, respectively.

We are grateful to Josh Angrist, Janet Currie, Michael Greenstone, Christopher Jenks, Alan Krueger, James Heckman, Susan Mayer, Derek Neal, Meredith Phillips, Barbara Schneider, and two anonymous referees for helpful comments and suggestions. Financial support was provided by the National Science Foundation to Fryer and Levitt. Correspondence can be addressed to Fryer at Harvard Society of Fellows, Littauer Center, Cambridge MA, 02138, or to Levitt at Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637. E-mail: rfryer@fas.harvard.edu, slevitt@midway.uchicago.edu.

[1] See Baughman and Dahlstrom (1968), Bracken, Sabers, and Insko (1987), Brooks-Gunn et al. (1993a, 1994, 1996), Coleman et al. (1966), Coley (2002), Hernstein and Murray (1994), Humphreys (1988), Jensen (1969, 1973), Kaufman and Kaufman (1983), Krohn and Lamp (1989), Naglieri (1986), Phillips et al. (1998), Phillips (2000), and Scarr (1981).

[2] To this effect, Jenks and Phillips (1998a) write, "Reducing the black-white test score gap would do more to promote racial equality than any other strategy that commands broad political support."

[3] On a test of general knowledge, a racial test score gap persists. On a subjective teacher assessment of general knowledge, however, there is no difference between blacks and whites in the fall of kindergarten.

same set of covariates accounts for much but not all of the Hispanic-white difference in test scores, but cannot explain the high test scores of Asians.

There are three leading explanations for why our results differ so sharply from earlier research such as Phillips et al. (1998a): (1) nonrandom sampling in the data sets used in earlier studies, (2) real gains by recent cohorts of blacks, and (3) better covariates in ECLS. Based on our analysis of the Children of the National Longitudinal Survey of Youth (CNLSY) data used by Phillips et al. (1998a), we conclude that real gains by recent cohorts of blacks are an important part of the explanation. The raw black-white test score gap for recent cohorts in CNLSY are comparable to those in ECLS, in sharp contrast to earlier cohorts in CNLSY. Real gains by blacks born in recent years would appear to be the leading explanation. We cannot, however, fully eliminate the racial test score gap among recent CLNSY cohorts. This is due in part to better covariates in ECLS. Even when nearly identical covariates are included, differences persist between ECLS and CNLSY.

Despite the fact that we see no difference in initial test scores for observationally equivalent black and white children when they enter kindergarten, their paths diverge once they are in school. Between the beginning of kindergarten and the end of first grade, black students lose 0.20 standard deviation (approximately 0.10 standard deviation each year) relative to white students with similar characteristics.[4] If the gap in test scores for these children continues to grow at the same rate, by fifth grade the black students will be 0.50 standard deviation behind their white counterparts—a gap similar in magnitude to that found in previous analyses (Jones, Burton, & Davenport, 1982; Phillips et al., 1998b; Phillips, 2000).

The leading explanation for the worse trajectory of black students in our sample is that they attend lower-quality schools. When we compare the changes in test scores over time for blacks and whites attending the same school, we find that black students lose only a third as much ground as they do relative to whites in the overall sample. This result suggests that differences in quality across schools attended by whites and blacks is likely to be an important part of the story. Interestingly, along traditionally considered dimensions of school quality (class size, teacher education, computer : student ratio, and so on), blacks and whites attend schools that are similar. On a wide range of nonstandard school inputs (including gang problems in school, percentage of students on free lunch, amount of loitering in front of school by nonstudents, amount of litter around the school, whether or not students need hall passes, and PTA funding), blacks do appear to be attending much worse schools even after controlling for individual characteristics.[5] Our story is incomplete, however, because the observable differences across schools do little to explain the widening black-white gap. This could be due to the coarseness of the school quality variables available in the ECLS.

We explore a range of other explanations as to why black children are losing ground, but find very little empirical support for these alternative theories. Black students do not appear to suffer bigger summer setbacks when school is not in session. The lower trajectories of black students is not simply an artifact of standardized testing. Subjective teacher assessments of student performance yield patterns similar to the test score data. Having a black teacher provides no benefit to black students compared to their white classmates, calling into question the possible role of either overt discrimination or low expectations for black children on the part of white teachers. Finally, adding proxies for behavioral problems does not alter our findings.

The structure of the paper is as follows. Section II provides a brief review of the literature. Section III describes and summarizes the data set. Section IV presents the basic results for incoming kindergartners, demonstrating that the black-white test score gap disappears once other confounding factors are allowed for. Section V documents the fact that a racial test score gap emerges during the school-age years, and Section VI analyzes the reasons for this divergence. Section VII concludes.

II. Background and Previous Literature

The Coleman report (Coleman et al., 1966) was the first national study to describe ethnic differences in academic achievement among children at various stages of schooling. It documented that substantial differences in educational achievement between blacks and whites not only existed at every grade level, but increased with student age. Since then, substantial effort has been devoted to understanding what variables account for the gap, as well as how and why the magnitude of the gap has changed over time.[6] A number of stylized facts have emerged. Socioeconomic status and the effects of poverty are important factors in explaining racial differences in educational achievement (Brooks-Gunn & Duncan, 1997; Mayer, 1997; Brooks-Gunn et al., 1994, 1995, 2000). Even after controlling for socioeconomic status in conventional regression analysis, a substantial gap still remains. That gap has generally been declining over time, although for high school students today, it is slightly larger than it was in the late 1980s (Grissmer, Flanagan, and Williamson, 1998; Hedges & Nowell, 1998; Humphreys, 1988). Finally, the

[4] Neither Hispanics nor Asians experience this widening test score gap over time. Indeed, Hispanic children systematically close the gap relative to whites, presumably because their initial scores are artificially low as a consequence of limited English proficiency among some Hispanic parents.

[5] This pattern is also consistent with self-selection of low-achieving whites into schools attended by blacks. Casting doubt on this alternative explanation is the fact that whites who go to school with blacks have baseline test scores upon entering kindergarten that are similar to those who are in all-white classes [Humphreys (1988) documents a similar finding among high school students]. When we eliminate from the sample whites who have no black children in their class (more than 60% of all white children fall into this category), we obtain similar results.

[6] In particular, Hernstein and Murray's controversial book, *The Bell Curve,* published in 1994, ignited interest in the subject by arguing that genetic differences are the primary explanation for the differences between blacks and whites in achievement test scores. For excellent summaries of the book, see Heckman (1995) and Goldberg and Manski (1995). Examples of the discussion that emerged include Devlin, Resnick, and Roeder (1998), Fraser (1995), and Kincheloe, Steinberg, and Gresson (1997).

gap in test scores between blacks and whites historically emerges before children enter kindergarten and tends to widen over time (Phillips et al., 1998b; Carneiro & Heckman, 2002).

III. The Data

The Early Childhood Longitudinal Study kindergarten cohort (ECLS-K) is a nationally representative sample of over 20,000 children entering kindergarten in 1998. Thus far, information on these children has been gathered at four separate points in time. The full sample was interviewed in the fall and spring of kindergarten and the spring of first grade. A random sample of one-fourth of the respondents were also interviewed in the fall of first grade. The sample will eventually be followed through fifth grade.[7] Roughly 1,000 schools are included in the sample, with an average of more than twenty children per school in the study. As a consequence, it is possible to conduct within-school analyses.

A wide range of data are gathered on the children in the study, as described in detail at the ECLS website http://nces.ed.gov/ecls. We utilize just a small subset of the available information in our baseline specifications (although we also show that similar results are obtained in a much more fully specified model). Students who are missing data on test scores, race, or age are dropped from our sample.

Summary statistics for the variables we use in our core specifications are displayed by race in table 1, with white referring solely to non-Hispanic whites.[8] Our primary outcome variables are math and reading standardized test scores.[9] Standardized tests were administered orally to the full sample in the fall of kindergarten and in the spring of first grade.[10] The reading test includes questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, and word recognition), vocabulary and comprehension, listening and reading comprehension, knowledge of the alphabet, phonetics, and so on. The math test evaluates number recognition, counting, comparing and ordering numbers, solving word problems, and interpreting picture graphs. The values reported in the table are item response theory (IRT) scores provided in ECLS-K, which we have transformed to have mean 0 and standard deviation 1 for the overall sample on each of the tests and time periods.[11] In all instances sample weights provided in ECLS-K are used.[12] White students on average score 0.274 standard deviation above the mean on the math exam in the fall of kindergarten, whereas black students perform 0.364 standard deviation below the mean on that test, yielding a black-white gap of 0.638 standard deviation. By the spring of first grade, that gap has increased to 0.728 standard deviation. The initial black-white gap in reading is smaller (0.401 standard deviation). Like the math gap, however, the reading gap widens substantially, to 0.529 standard deviation, by the end of first grade. It is worth noting that the black-white gaps are substantially smaller than those observed in earlier data sets for children of the same age. For instance, Phillips et al. (1998a) report a raw black-white test score gap of over 1 standard deviation in reading using the 1980–1987 cohorts of CNLSY and the 1984–1985 cohorts of the Infant Health and Development Program (IHDP) data set.

A second outcome measure that we analyze is subjective teacher assessments of a child's math and reading achievement. Teachers were asked to answer 20 questions about the child's academic performance, ranking them on a scale from "not yet" to "proficient." These answers were then transformed into IRT scores. As was done with test scores, these subjective assessments have been renormalized to have mean 0 and standard deviation 1. The patterns in the teacher assessments mirror those in the test score data: black and Hispanic students start out substantially below whites, and black students lose ground over the first two years of school, whereas Hispanics maintain their position relative to whites. The most notable difference between the test scores and teacher assessments is that Asian students are rated at or below the level of white students in the fall of kindergarten on the teacher assessments, but then gain relative to whites over time.

[7] In addition, there is an ECLS birth cohort that tracks a nationally representative sample of over 15,000 children born in 2001 through the first grade.

[8] There are also a small number of children in the data whose racial status is classified as "other." These include Hawaiian, mixed race, and Native American students. Such students are included in our regressions, but not shown in the summary statistics table.

[9] These tests were developed especially for the ECLS, but are based on existing instruments, including Children's Cognitive Battery (CCB), Peabody Individual Achievement Test—Revised (PIAT-R); Peabody Picture Vocabulary Test—3 (PPVT-3); Primary Test of Cognitive Skills (PTCS); and Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R). Students are administered the test questions orally, as it is not assumed that they know how to read. A *general knowledge* exam was also administered. It is designed to capture "children's knowledge and understanding of the social, physical, and natural world and their ability to draw inferences and comprehend implications." No further information is available on the precise content of the general knowledge exam questions or skills tested. We limit the analysis to math and reading scores, primarily because of the comparability of these test scores with past research in the area. In addition, there appear to be some peculiarities in the results of the general knowledge exam. For instance, Asians score well above other groups on math and reading, but do extremely poorly on the general knowledge exam. Also, black students do extremely poorly on the general knowledge exam, even though teachers rate them only slightly behind whites in this area on the subjective teacher evaluations. Most of our results also appear in the general knowledge scores, and we note the instances where differences arise.

[10] The tests were also given in the spring of kindergarten, but we limit our focus to the endpoints of the available data. The kindergarten spring test results are in all cases consistent with the results presented in this paper.

[11] Because children were asked different questions depending on the answers they provided to the initial questions on the test, IRT-adjusted scores are preferable to simple test score measures reflecting the number of correct answers a child provided. For more detail on the process used to generate the IRT scores, see Chapter 3 of the ECLS-K User's Guide (NCES, 2002). Our results are not sensitive to normalizing the IRT scores to have mean 0 and standard deviation 1.

[12] Because of the complex manner in which the ECLS-K sample is drawn, different weights are suggested by the providers of the data depending upon the set of variables used. We utilize the weights recommended for making longitudinal comparisons. None of our findings are sensitive to other choices of weights, or to not weighting at all.

TABLE 1.—SUMMARY STATISTICS BY RACE: STUDENT CHARACTERISTICS

Variable	Full Sample	White	Black	Hispanic	Asian
Test Scores:					
Fall kindergarten math	.026	.274	−.364	−.448	.424
	(1.038)	(1.073)	(.765)	(.898)	(1.275)
Spring first grade math	.013	.249	−.479	−.314	.272
	(1.092)	(.984)	(.977)	(1.040)	(1.089)
Fall kindergarten reading	.014	.147	−.255	−.280	.481
	(1.010)	(1.073)	(.892)	(.944)	(1.459)
Spring first grade reading	.033	.175	−.354	−.134	.472
	(1.123)	(1.073)	(1.062)	(1.062)	(1.088)
Subjective teacher assessments:					
Fall kindergarten math	.071	.249	−.178	−.286	.129
	(1.115)	(1.084)	(1.001)	(.988)	(1.269)
Spring first grade math	−.005	.130	−.355	−.163	.234
	(1.115)	(1.084)	(1.101)	(1.102)	(.983)
Fall kindergarten reading	.081	.267	−.114	−.354	.063
	(1.041)	(1.050)	(1.025)	(1.043)	(1.212)
Spring first grade reading	.006	.118	−.252	−.146	.206
	(1.145)	(1.050)	(1.177)	(1.127)	(1.095)
Race:					
White	.578	1.000	.000	.000	.000
	(.550)				
Black	.161	.000	1.000	.000	.000
	(.390)				
Hispanic	.189	.000	.000	1.000	.000
	(.390)				
Asian	.028	.000	.000	.000	1.000
	(.130)				
Other	.045	.000	.000	.000	.000
	(.260)				
Other controls:					
Female	.486	.481	.495	.494	.503
	(.520)	(.594)	(.536)	(.542)	(.591)
Age (in months), fall of kindergarten	67.061	67.402	66.877	66.413	65.927
	(4.811)	(4.752)	(4.776)	(4.773)	(4.639)
SES composite measure	−.018	.202	−.359	−.423	.312
	(.910)	(.792)	(.780)	(.759)	(1.034)
Number of children's books in the home	71.946	93.121	39.014	40.849	49.333
	(64.551)	(64.792)	(41.986)	(47.752)	(55.859)
Mother's age at time of first birth	23.342	24.579	20.548	21.762	25.807
	(6.052)	(5.916)	(5.139)	(5.256)	(6.330)
Child's birth weight (in ounces)	117.827	120.256	110.315	117.091	112.667
	(23.847)	(22.942)	(25.373)	(23.048)	(22.498)
WIC participation	.475	.332	.772	.652	.313
	(.517)	(.493)	(.481)	(.539)	(.529)
Frequency of missing values:					
Missing number of children's books in the home	.010	.010	.008	.007	.017
	(.130)	(.099)	(.097)	(.108)	(.148)
Missing WIC	.012	.007	.029	.014	.010
	(.130)	(.099)	(.195)	(.108)	(.118)
Missing mother's age	.020	.008	.049	.022	.045
	(.130)	(.099)	(.244)	(.163)	(.207)
Missing birth weight	.128	.100	.150	.170	.246
	(.390)	(.297)	(.390)	(.434)	(.502)

NOTES: The entries are means and standard deviations of student-level data for those students in ECLS-K who do not have missing values for test scores, race, or age. Test scores are IRT scores, normalized to have a mean of 0 and a standard deviation of 1 in the full, unweighted sample. Subjective teacher assessments have also been normalized in this manner. The category white includes only non-Hispanic whites. Precise definitions of the variables are provided in the data appendix. The SES composite measure incorporates information on parental education, occupational status, and family income. The SES measure ranges from −4.75 to 2.75 in the sample, with larger numbers indicating higher socioeconomic status. The total number of students in the sample who receive a positive weight in the estimation is 13,290. The bottom panel of the table reports the frequency of missing values for the covariates. In all cases, sample weights provided with ECLS are used in the calculations.

The remainder of table 1 presents summary statistics for the other variables used in the analysis. In contrast to the test score variables, for which we have observations at multiple points in time, most of the control variables either are collected only once (typically kindergarten fall, but in some cases kindergarten spring), or exhibit little variation over time for individual students. The most important of these covariates is a composite measure of socioeconomic status constructed by the researchers conducting the ECLS survey. The components used in the SES measure are parental education, parental occupational status, and household income. Other variables included as controls are gender, child's age at the time of enrollment in kindergarten, WIC participation (a nutrition program aimed at low income

TABLE 2.—THE ESTIMATED BLACK-WHITE TEST SCORE GAP IN FALL OF KINDERGARTEN

	Math					Reading				
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	−.638	−.368	−.238	−.094	−.102	−.401	−.134	−.006	.117	.093
	(.022)	(.022)	(.023)	(.023)	(.026)	(.024)	(.025)	(.026)	(.025)	(.030)
Hispanic	−.722	−.429	−.302	−.203	−.171	−.427	−.223	−.137	−.064	−.076
	(.022)	(.023)	(.024)	(.022)	(.028)	(.027)	(.026)	(.026)	(.025)	(.029)
Asian	.150	.070	.190	.265	.274	.335	.256	.371	.409	.375
	(.056)	(.051)	(.051)	(.048)	(.050)	(.064)	(.059)	(.059)	(.058)	(.060)
Other race	−.503	−.329	−.253	−.158	−.113	−.401	−.230	−.155	−.072	−.014
	(.041)	(.037)	(.036)	(.035)	(.035)	(.044)	(.040)	(.040)	(.038)	(.039)
Socioeconomic status composite measure		.456	.389	.302	.072		.451	.393	.299	.092
	—	(.014)	(.014)	(.014)	(.024)	—	(.014)	(.015)	(.015)	(.023)
Number of children's books			.007	.006	.005			.007	.006	.004
	—	—	(.001)	(.001)	(.001)	—	—	(.001)	(.001)	(.001)
(Number of children's books)^2 (×1000)			−.023	−.020	−.027			−.025	−.021	−.017
	—	—	(.003)	(.002)	(.016)	—	—	(.003)	(.003)	(.017)
Female				.010	.00				.159	.153
	—	—	—	(.015)	(.015)	—	—	—	(.017)	(.016)
Age at kindergarten fall (in months)				.056	−2.680				.041	−2.409
	—	—	—	(.002)	(.542)	—	—	—	(.002)	(.483)
Birth weight (ounces) (× 10)				.029	.030				.019	.022
	—	—	—	(.004)	(.004)	—	—	—	(.004)	(.004)
Teenage mother at time of first birth				−.109	−.029				−.144	−.069
	—	—	—	(.018)	(.021)	—	—	—	(.020)	(.022)
Mother at least 30 at time of first birth				.182	.11				.226	.155
	—	—	—	(.025)	(.028)	—	—	—	(.027)	(.030)
WIC participant				−.211	−.120				−.184	−.104
	—	—	—	(.019)	(.020)	—	—	—	(.021)	(.021)
R^2	0.108	0.223	0.239	0.317	0.354	0.045	0.16	0.175	0.233	0.279
Number of observations			13,290					12,601		
Full set of covariates included in regression?	N	N	N	N	Y	N	N	N	N	Y

NOTES: The dependent variable is the math or reading test score in the fall of kindergarten. Test scores are IRT scores, 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 in a student. Standard errors 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 students with missing values on each covariate are also included in the regressions. In addition, columns 5 and 10 report only a subset of the coefficients from regressions with 98 covariates included in the specification. The full results for columns 5 and 10 are reported in appendix table A1. Note that the specifications in columns 5 and 10 include age and age squared; that is why the coefficient on age changes so dramatically relative to other columns in the table.

mothers and children), mother's age at first birth, birth weight, and the number of children's books in the home.[13]

There are substantial differences across races on many of these variables. Black children in the sample are growing up under circumstances likely to be less conducive to academic achievement than white children: lower socioeconomic status, fewer children's books in the home, and so on. Hispanics are also worse off than whites on average. For Asians, the patterns are more mixed. Though this may seem an odd set of covariates to include, our rationale is that the results we obtain with this small set of variables mirrors the findings when we include an exhaustive set of over 100 controls.[14]

We caution against a causal interpretation of the coefficients on the covariates, which we view as proxies for a broad set of environmental and behavioral factors. In view of past research that has had great difficulty making the black-white test score gap disappear, we focus on the results from these very sparse regressions to highlight the fact that the sharp differences between our results and earlier studies are not primarily a consequence of the availability of different covariates in the ECLS.

IV. Estimating Racial Test Score Gaps for Incoming Kindergartners

Table 2 presents a series of estimates of the racial test score gap for the tests taken in the fall of kindergarten. The specifications estimated are of the form

$$TESTSCORE_i = RACE_i'\Gamma + X_i'\Theta + \varepsilon_i, \qquad (1)$$

where i indexes students. A full set of race dummies are included in the regression, with white as the omitted category. Consequently, the coefficients on race capture the gap between the named racial category and whites. Our primary emphasis is on the black-white test score gap. The vector of other covariates included in the specification, denoted X_i, varies across columns in table 2. As one moves to the right in the table, the set of covariates steadily grows. In all instances, the estimation is done using weighted least squares, with weights corresponding to the sampling weights provided in the data set.

The first and sixth columns of table 2 present the differences in means, not including any covariates. These results

[13] A more detailed description of each of the variables used is provided in the appendix.

[14] We also present the results using a fuller set of controls for completeness.

simply reflect the raw test score gaps reported in table 1. The next specification adds the composite indicator of socioeconomic status constructed by the ECLS survey administrators. Socioeconomic status is an important predictor of incoming test scores, carrying a t-statistic over 40. A 1-standard-deviation increase in the SES variable is associated with a 0.41 increase in both math and reading test scores. Controlling for socioeconomic status substantially reduces the estimated racial gaps in test scores (see also Coley, 2002). The black-white gap in math falls by more than 40%; the reading gap is reduced by more than two-thirds. The changes in the other race coefficients are not as large, but in every instance the estimated gaps shrink, and R^2 increases substantially.

The next set of specifications adds the number of children's books in the child's home, the square of that variable, and an indicator variable equal to 1 if the number of books takes on a missing value for that student. The number of books is strongly positively associated with high kindergarten test scores on both math and reading.[15] Evaluated at the mean, a 1-standard-deviation increase in the number of books (from 72 to 137) is associated with increases of 0.143 and 0.115 in math and reading, respectively. This variable seems to serve as a useful proxy for capturing the conduciveness of the home environment to academic success. Including number of books reduces the black-white gap on math to less than one-fourth of a standard deviation and completely eliminates the gap in reading. The gap for Hispanics also shrinks. The Asian-white gap, however, becomes even larger than the raw gap when number of books is added to the regression.

Columns 4 and 9 add controls for gender, age, birth weight, indicator variables for having a mother whose first birth came when she was a teenager or over 30 (the omitted category is having a first birth in one's twenties), and WIC participation. These covariates generally enter with the expected sign. Older children, those with higher birth weights, and those with older mothers at the time of first birth all score better. Children on WIC do worse on the tests, suggesting that this variable is not capturing any real benefits the program might provide, but rather, the fact that eligibility for WIC is a proxy for growing up poor, which the SES variable is not adequately capturing. Adding these variables to the specification further improves the test scores of blacks and Hispanics. In fact, the estimates suggest that, controlling for other factors, black children actually score slightly better than whites in reading, and only slightly worse in math. We do not have a compelling explanation why there is a difference between reading and math achievement. Only a small gap persists for Hispanics. The advantage enjoyed by Asians becomes even greater. R^2 increases substantially relative to the previous specification.

The final specifications in table 2 (columns 5 and 10) include an exhaustive set of roughly 100 covariates capturing city size, neighborhood characteristics, region of the country, parental education, parental income, parental occupational status, family size and structure, whether the mother worked, type of preschool program participation, whether English is spoken at home, and the extent of parental involvement in a child's life and school. We report only a subset of the covariates in table 2; full results are presented in appendix table A1. Almost all of the controls enter in the predicted direction and with coefficients of plausible magnitude. Interestingly, none of the coefficients on race change appreciably. Only a few of the parameters on the controls included in the parsimonious specifications are greatly affected either, and these are easily explained. The socioeconomic status coefficient shrinks because the full set of covariates includes variables that go into the construction of the composite indicator such as parent's income and occupational status. The coefficient on age becomes highly negative because an age squared term (which is positive and significant) is included in the full specification. The inclusion of these additional variables does little to improve the fit of the model.

Table 3 explores the sensitivity of the estimated racial gaps in test scores across a wide variety of alternative specifications and subsamples of the data. We report only the race coefficients and associated standard errors. The top row of the table presents the baseline results using a full sample and our parsimonious set of controls (corresponding to columns 4 and 9 of table 2). Weighting all of the observations equally in the regressions leaves the black-white gaps in math and reading virtually unchanged. Employing an alternative test-score measure (T-scores, which are norm-referenced measurements of achievement) has very little impact on the results.

One might be concerned that restricting all the coefficient estimates to be identical across the entire sample may yield misleading results. Regressions on a common support (for example, only on single mothers, in only one region of the country, or only in rural areas) provide one means of addressing this concern. Almost every subset of the data examined yields results roughly similar to those for the overall sample. There is some slight evidence that black females do better relative to whites than do black males. The results appear to be quite consistent across quintiles of the socioeconomic status distribution. Due in part to relatively imprecise estimates, the equality of black and white test scores on math and reading tests can rarely be rejected for any of the quintiles. Rural blacks do somewhat worse relative to whites than those in central cities. Blacks in

[15] The marginal benefit associated with one additional book decreases as more books are added. Beyond roughly 150 books, it turns negative. Only 16% of the sample lies above this cutoff point.

TABLE 3.—SENSITIVITY ANALYSIS AND EXTENSIONS OF THE BASIC MODEL FOR FALL KINDERGARTEN TEST SCORES

Specification	Coefficient on Black for: Math	Coefficient on Black for: Reading	Coefficient on Hispanic for: Math	Coefficient on Hispanic for: Reading	Coefficient on Asian for: Math	Coefficient on Asian for: Reading
Baseline	−.094 (.023)	.117 (.025)	−.203 (.022)	−.064 (.025)	.265 (.048)	.409 (.058)
Unweighted	−.100 (.023)	.092 (.024)	−.206 (.021)	−.057 (.024)	.285 (.034)	.387 (.035)
Other test score measures:						
T-scores	−.050 (.024)	.141 (.030)	−.057 (.022)	.065 (.028)	.176 (.040)	.298 (.048)
By gender						
Males	−.126 (.034)	.093 (.037)	−.224 (.032)	−.095 (.035)	.338 (.078)	.385 (.087)
Females	−.058 (.030)	.147 (.035)	−.181 (.031)	−.035 (.036)	.203 (.059)	.433 (.077)
By SES quintile:						
Bottom	−.092 (.044)	−.005 (.041)	−.202 (.044)	−.133 (.045)	.328 (.143)	.043 (.111)
Second	−.088 (.045)	.091 (.049)	−.179 (.046)	−.090 (.047)	.044 (.106)	−.001 (.090)
Third	−.097 (.049)	.068 (.045)	−.242 (.046)	−.106 (.051)	.249 (.121)	.351 (.167)
Fourth	−.082 (.058)	.292 (.077)	−.100 (.056)	.030 (.057)	.207 (.088)	.396 (.115)
Top	−.169 (.080)	.068 (.085)	−.323 (.078)	−.113 (.094)	.404 (.087)	.724 (.102)
By family structure:						
Single mother	−.087 (.043)	.070 (.043)	−.197 (.048)	−.119 (.047)	.086 (.149)	.114 (.144)
Two biological parents	−.127 (.034)	.141 (.042)	−.176 (.029)	−.033 (.033)	.291 (.054)	.456 (.064)
Teen mother at 1st birth	−.101 (.036)	.014 (.033)	−.199 (.036)	−.127 (.038)	.170 (.105)	.251 (.114)
Teen mother at child's birth	−.062 (.046)	−.021 (.043)	−.196 (.045)	−.105 (.052)	.279 (.141)	.281 (.135)
By region:						
Northeast	−.087 (.060)	.129 (.076)	−.159 (.054)	−.030 (.060)	.305 (.124)	.483 (.156)
Midwest	.004 (.053)	.093 (.057)	−.140 (.064)	−.031 (.061)	.337 (.119)	.562 (.133)
South	−.153 (.032)	.051 (.033)	−.217 (.040)	−.119 (.048)	.154 (.104)	.368 (.111)
West	.098 (.077)	.362 (.095)	−.200 (.044)	−.001 (.048)	.269 (.071)	.353 (.088)
By location type:						
Central city	−.110 (.035)	.147 (.040)	−.235 (.033)	−.073 (.037)	.271 (.061)	.439 (.075)
Suburban	−.135 (.039)	.030 (.041)	−.261 (.041)	−.145 (.042)	.146 (.102)	.310 (.119)
Rural	−.184 (.048)	−.032 (.050)	−.253 (.062)	−.124 (.072)	.255 (.130)	.126 (.102)
By school type:						
Public	−.106 (.024)	.098 (.027)	−.214 (.024)	−.081 (.027)	.260 (.051)	.392 (.064)
Private	.022 (.070)	.281 (.074)	−.152 (.058)	.015 (.066)	.296 (.135)	.479 (.137)
School >80% black	.053 (.269)	−.016 (.215)	−.084 (.298)	.057 (.273)	.285 (.382)	.788 (.641)
School >80% white	−.105 (.047)	.059 (.053)	−.186 (.025)	−.061 (.028)	.288 (.054)	.436 (.065)

NOTES: Specifications in this table are variations on those reported in columns 4 and 9 of table 2. Only the race coefficients are reported. The top row simply reproduces the baseline results in columns 4 and 9 of table 2. The remaining rows correspond to different weights, test score measures, or particular subsets of the data. For further details of the baseline specification, see the notes to table 2.

private schools appear to do especially well, consistent with Neal (1997) and Grogger and Neal (2000).[16]

The results presented in tables 2 and 3 maintain the assumption that children of different races are equally responsive to changes in covariates. Cross-race differences in coefficients are potentially important because they affect the interpretation of the racial test score gap estimates in the preceding tables. Black children experience worse environments on average. If black children do not derive as much benefit from improvements in socioeconomic status, number of children's books, higher birth weight, and so on, then our earlier results will overstate the convergence in black-white test scores.

Table 4 presents within-race estimates of our basic specifications to determine how large this bias may be. Columns 1 and 6 replicate the coefficient estimates from the full sample. The remaining columns present results within a specific race category. The black children in our sample are less responsive to changes in socioeconomic status than whites: a 1-standard-deviation improvement in socioeconomic status for a black child is associated with a 0.176-standard-deviation increase in math scores, compared to 0.316 for a white child. For most of the other covariates, however, the white and black coefficients are similar. Using the coefficients in columns 3 and 8, a black child that had the characteristics of the average white child in the sample would be estimated to score −0.21 standard deviations below that white child on math and be almost exactly even in reading.[17] Thus, to the extent that public policies are designed to improve the environments experienced by black children, our baseline estimates may slightly overstate the ground that would be gained by blacks. This logic also holds for Hispanics, but not for Asians, who seem to be more responsive to environmental influences.

The fact that the black-white test score gap essentially disappears with the inclusion of sufficient controls in ECLS is a very striking result, in that in past research a substantial

[16] We have also experimented with limiting the sample to the set of children for whom there is substantial overlap across races in background characteristics. More specifically, we ran probits with an indicator variable for black as the dependent variable and the full set of covariates as predictors. When we drop from the sample the roughly 30% of students whose predicted probability of being black is less than 10% or greater than 90%, the black-white gap on math rises slightly and the reading gap becomes closer to 0.

[17] Using the coefficients in columns 2 and 7, a white student that had the characteristics of the average black child would be expected to score 0.07 (0.13) standard deviation above (below) that average black child on math (reading). These results are almost identical to the predictions from our baseline specification, because the coefficients from the white regression are quite similar to those from the full sample.

TABLE 4.—ESTIMATES OF THE RESPONSIVENESS OF TEST SCORES TO COVARIATES BY RACE

	Math					Reading				
Variable	Full Sample	White	Black	Asian	Hispanic	Full Sample	White	Black	Asian	Hispanic
Black	−.094	—	—	—	—	.117	—	—	—	—
	(.023)					(.025)				
Hispanic	−.203	—	—	—	—	−.064	—	—	—	—
	(.022)					(.025)				
Asian	.265	—	—	—	—	.409	—	—	—	—
	(.048)					(.058)				
Other race	−.158	—	—	—	—	−.072	—	—	—	—
	(.035)					(.038)				
Socioeconomic status composite	.302	.347	.193	.400	.212	.299	.326	.217	.556	.203
measure	(.014)	(.020)	(.030)	(.071)	(.031)	(.015)	(.020)	(.033)	(.080)	(.038)
Number of books	.006	.005	.005	.011	.009	.006	.006	.004	.012	.007
	(.001)	(.001)	(.001)	(.003)	(.001)	(.001)	(.001)	(.001)	(.004)	(.002)
(Number of books)^2 × 1000	−.020	−.017	−.017	−.040	−.032	−.021	−.020	−.012	−.044	−.025
	(.002)	(.003)	(.007)	(.015)	(.006)	(.003)	(.003)	(.010)	(.018)	(.008)
Female	.010	.003	.058	−.151	.029	.159	.170	.146	.152	.163
	(.015)	(.021)	(.033)	(.096)	(.032)	(.017)	(.022)	(.038)	(.113)	(.043)
Age at kindergarten fall (in months)	.056	.061	.045	.074	.051	.041	.044	.035	.052	.035
	(.002)	(.003)	(.004)	(.015)	(.004)	(.002)	(.003)	(.004)	(.018)	(.005)
Birth weight (ounces) × 10	.029	.037	.023	.084	.002	.019	.024	.021	.088	−.012
	(.004)	(.005)	(.007)	(.032)	(.009)	(.004)	(.005)	(.009)	(.045)	(.010)
Teenage mother at time of first birth	−.109	−.126	−.127	−.113	−.073	−.144	−.117	−.219	−.022	−.129
	(.018)	(.030)	(.036)	(.123)	(.033)	(.020)	(.029)	(.041)	(.144)	(.044)
Mother at least 30 at time of first birth	.182	.174	.106	.233	.178	.226	.206	.231	.132	.348
	(.025)	(.030)	(.071)	(.118)	(.066)	(.027)	(.031)	(.095)	(.145)	(.085)
WIC participant	−.211	−.204	−.173	−.087	−.198	−.184	−.177	−.147	−.203	−.184
	(.019)	(.027)	(.047)	(.108)	(.039)	(.021)	(.027)	(.056)	(.124)	(.049)
R^2	0.317	.227	.194	.300	.301	0.233	.189	.181	.285	.234
Number of obs.	13290	7,999	1,806	537	2,234	12601	8,000	1,804	536	1,546

NOTES: The dependent variable is the math or reading test score in the fall of kindergarten. The first and sixth columns simply replicate our baseline results from columns 4 and 9 of table 2. The other columns report coefficient estimates from specifications identical to that of the baseline, but estimated separately by race of the student. There is no within-race variation in the race variables, so these variables are omitted from the within-race regressions. See the notes to table 2 for further details of the estimation.

gap has persisted, regardless of the age of the individuals, the particular tests, or the covariates included (see for example Hernstein & Murray, 1994; Neal & Johnson, 1996; Phillips et al., 1998a).[18] The most direct comparison for our research among previous studies is Phillips et al. (1998a), which looks at test outcomes for kindergartners in the early cohorts of CNLSY. Although Phillips et al. (1998a) have the greatest success among earlier studies in explaining the racial differences in reading (they reduce the gap by two-thirds with their covariates), their raw gap is so large compared to ECLS that the *residual* gap in that paper is almost as large as the *raw* gap in ECLS.

Why our results differ so sharply from previous research, and Phillips et al. (1998a) in particular, is a question of critical importance. There are three leading explanations for the divergence: (1) the sample of births included in CNLSY, especially in the early years, may be nonrepresentative, (2) better covariates are available in ECLS, and (3) blacks born into recent cohorts have made real gains relative to blacks born a decade earlier. The first two explanations appear to play only a small role empirically. Although it is true that the sample of births in early cohorts of CNLSY analyzed by Phillips et al. (1998a) is heavily skewed toward teenage mothers because of the way the sample is generated (that is, by births to those included in NLSY), the nonrandom sampling does not seem to provide the explanation for the differing results. When we restrict our ECLS sample to children born to teen mothers, our results are virtually unchanged.[19] When we try to estimate specifications in ECLS using only variables that are available in CNLSY, blacks do somewhat worse than in our baseline sample (a gap of −0.183 on math and 0.034 on reading), but this is nothing like the residual gap of −0.67 on reading in Phillips et al. (1998a).

Real gains by blacks in recent cohorts, in contrast, do appear to be an important part of the divergence between our results and past research. Limiting the CNLSY to

[18] The exceptions we are aware of in which the black-white test score gap has been made to disappear are Crane (1994), Li and Poirier (2001), and Carneiro and Heckman (2002). Li and Poirier (2001), using a Bayesian structural model, find no systematic differences between blacks and whites using the NLSY. Hernstein and Murray (1994) and Phillips et al. (1998a), using different methods on the same data, find large gaps still persists. Using CNLSY, Crane (1998) and Carneiro and Heckman (2002) find that on some tests, racial gaps disappear with controls, although large gaps remain on other tests designed to capture similar sets of skills. It is important to note that on the test of general knowledge in ECLS, the black-white gap does not fully disappear. Black students test almost 1 full standard deviation behind whites in a raw comparison of means. That gap falls to 0.3 when controls are included. On the subjective teacher assessments, the raw gap in general knowledge between blacks and whites is much smaller (0.25 standard deviation) and does shrink almost to 0 with the inclusion of controls.

[19] Our results are also unchanged when we limit our ECLS sample to low-birth-weight babies, who are oversampled in IHDP, another data set analyzed by Phillips et al. (1998a).

TABLE 5.—THE EVOLUTION OF TEST SCORE GAPS BY RACE AS CHILDREN AGE

Variable	Math			Reading		
	Fall Kindergarten	Spring Kindergarten	Spring First Grade	Fall Kindergarten	Spring Kindergarten	Spring First Grade
Black	−.094 (.023)	−.201 (.025)	−.250 (.028)	.117 (.025)	−.009 (.027)	−.071 (.029)
Hispanic	−.203 (.022)	−.187 (.024)	−.120 (.026)	−.064 (.025)	−.005 (.027)	.001 (.029)
Asian	.265 (.048)	.221 (.049)	.115 (.044)	.409 (.058)	.434 (.054)	.345 (.045)
Other race	−.158 (.035)	−.166 (.039)	−.195 (.042)	−.072 (.038)	−.099 (.039)	−.163 (.042)
SES composite measure	.302 (.014)	.284 (.014)	.263 (.014)	.299 (.015)	.280 (.015)	.284 (.014)
Number of books	.006 (.001)	.006 (.001)	.005 (.001)	.006 (.001)	.005 (.001)	.006 (.001)
(Number of books)^2 ×1000	.020 (.002)	−.019 (.003)	−.019 (.003)	−.021 (.003)	−.020 (.003)	−.022 (.003)
Female	.010 (.015)	.003 (.016)	−.033 (.017)	.159 (.017)	.195 (.017)	.216 (.017)
Age at kindergarten fall (in months)	.056 (.002)	.051 (.002)	.036 (.002)	.041 (.002)	.034 (.002)	.021 (.002)
Birth weight (ounces) ×100	.029 (.004)	.003 (.000)	.029 (.004)	.019 (.004)	.002 (.000)	.024 (.005)
Teenage mother at time of first birth	−.109 (.018)	−.112 (.021)	−.111 (.022)	−.144 (.020)	−.138 (.021)	−.131 (.024)
Mother in 30s at time of first birth	.182 (.025)	.127 (.024)	.093 (.022)	.226 (.027)	.158 (.025)	.085 (.024)
WIC participant	−.211 (.019)	−.195 (.020)	−.201 (.021)	−.184 (.021)	−.152 (.02)	−.182 (.022)
R^2	0.317	0.282	0.240	0.233	0.197	0.194
Number of obs.	13,290	13,290	13,290	12,601	12,601	12,601

NOTES: The dependent variable is fall kindergarten test scores in columns 1 and 3 and spring first grade test scores in columns 2 and 4. All specifications include the parsimonious set of controls corresponding to columns 4 and 9 of table 2. Test scores are IRT scores, 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 student. Standard errors 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 students with missing values on each covariate are also included in the regressions.

cohorts born in the same years as the ECLS sample, the raw test score gaps in the CNLSY are nearly half as large as in earlier cohorts of CNLSY used by Phillips et al. (1998a) and are remarkably close to those found in the ECLS. On the math skills test, the raw gaps are 0.638 and 0.665 respectively in ECLS and CNLSY. For reading, the gap is 0.401 in ECLS and 0.540 in the CNLSY. Real gains by blacks in recent years could explain this result. Interestingly, however, using the same set of controls that yield math and reading gaps in ECLS of −0.183 and 0.034 respectively, in recent cohorts of the CNLSY the estimated black-white residual gaps are −0.500 and −0.41 on math and reading. Thus, although the raw gaps are similar in ECLS and recent cohorts of CNLSY, larger residual gaps remain in CNLSY, for reasons we cannot explain.

V. The Evolution of the Racial Test Score Gaps As Children Age

The results of the previous section demonstrate that although black test scores lag whites by a large margin, the inclusion of a small number of covariates eliminates any systematic differences in the math and reading performance of whites and blacks entering kindergarten. Hispanics somewhat lag whites, and Asians exceed all of the other races. In this section, we explore how those racial gaps change over time.

In terms of raw test scores, simple calculations based on the numbers in table 1 show that black students lose some ground relative to whites between the fall of kindergarten and the spring of first grade: 0.090 standard deviation on math and 0.128 standard deviation on reading. Table 5 presents regression results for those two time periods. We report results only from our parsimonious regression specification; similar racial gaps emerge when the exhaustive set of covariates is included. Controlling for other factors in the regressions, black students appear to lose much more ground than they do in the raw means: −0.156 standard deviation in math, and −0.188 standard deviation in reading.[20] If black students in the sample continue to lose ground through ninth grade at the rate experienced in the first two years of school, they will lag white students on average by a full standard deviation in raw math and reading scores and over two-thirds of a standard deviation in math even after controlling for observable characteristics (substantially smaller for reading). Raw gaps of that magnitude would be similar to those found in previous studies of high-school-age children (Grissmer et al., 1998; Hedges & Nowell, 1998; Humphreys, 1988; Phillips et al., 1998a; Phillips, 2000).

In striking contrast to the black-white gap, Hispanics show gains relative to whites between the beginning of kindergarten and the end of first grade. Asians lose roughly as much ground as blacks in math (although they start ahead of whites) and also fall slightly in reading. Thus, black students are not only losing ground relative to whites, but even more so relative to Hispanics, and somewhat less relative to Asians.

VI. Why are Black Students Losing Ground in the First Two Years of School?

Why black students fare worse in the first two years of school is a question of paramount importance, for two reasons. First, knowing the source of the divergence may aid in developing public policies to alleviate the problem.

[20] Similar results (not shown in the table) are obtained when we include the full set of nearly 100 covariates. In those specifications, black students lose 0.136 standard deviation on math and 0.109 standard deviation on reading. Including the fall kindergarten test score as a covariate predicting the spring first grade test score also has little impact on the results: black students lose 0.192 (0.140) standard deviation in math (reading).

TABLE 6.—DOES DIFFERENTIAL SCHOOL QUALITY EXPLAIN BLACK STUDENTS LOSING GROUND?

	Full Sample of Students			Excluding Students Attending All White Schools					
Subject	(1) Fall Kindergarten	(2) Spring First Grade	(3) Difference (2) − (1)	(4) Fall Kindergarten	(5) Spring First Grade	(6) Difference (5) − (4)	(7) Fall Kindergarten	(8) Spring First Grade	(9) Difference (8) − (7)
Math	−.094	−.250	−.156	−.136	−.261	−.125	−.175	−.222	−.047
	(.023)	(.028)	(.036)	(.028)	(.034)	(.044)	(.034)	(.040)	(.052)
Reading	.117	−.071	−.188	.072	−.084	−.156	−.007	−.057	−.05
	(.025)	(.029)	(.038)	(.030)	(.035)	(.046)	(.038)	(.042)	(.057)
Include school fixed effects in regression?	N	N	N	N	N	N	Y	Y	Y
Number of obs.		13,290				6,532			

A comparison of cross-school and within-school estimates of the test score trajectory by race. Values reported are of the coefficient on the variable black.

NOTES: Entries are estimates of the black-white test score gap, controlling for the parsimonious set of regressors. Columns 3, 6, and 9 represent the estimated change in the gap between kindergarten fall and first grade spring. The first three columns include all students. The remaining columns restrict the data set to schools that had students of different races included in the ECLS-K sample. The final three columns include school fixed effects. Estimation is done using weighted least squares, using sample weights provided in the data set.

Second, determining the explanation for the widening gap will help to determine whether the simple linear extrapolation over the academic career is a plausible conjecture.

There are a number of plausible explanations as to why the racial gap in test scores grows as children age:

1. Black children attend lower quality schools on average.
2. The importance of parental and environmental contributions may grow over time. Because black children are on average disadvantaged in this regard, they fall behind.
3. Because of worse home and neighborhood environments, black students suffer worse summer setbacks when school is not in session.
4. The results are an artifact of the particular standardized tests used or of poor measurement of a child's environment, rather than representing true losses.
5. Something about the interaction between black students and schools interferes with the learning process. Such factors might include discrimination or low expectations on the part of teachers toward black students, systematic differences in self-control or socialization across children of different races, and the like.
6. The fall kindergarten test scores are measuring a different set of skills than the later tests, and the gap between whites and blacks is greater on the set of skills measured later.

We address each of these hypotheses in turn.

A. *Are Black Students Losing Ground Because They Attend Worse Schools?*

There is substantial racial segregation in school attendance in the United States. In our data, which sample roughly 20 children each from approximately 1,000 schools, in 35% of those schools there is not a single black child in the sample.[21] The mean black student in our sample attends a school that is 59% black and 8% Hispanic. In contrast, the typical white student goes to a school that is only 6% black and 5% Hispanic. Given that blacks and whites have little overlap in the schools they attend, differences in school quality are plausible explanations for why black students are losing ground.[22]

Because our data set has many individuals from each school included in the sampling frame, school fixed effects can be included in the estimation. With school fixed effects, the estimated black-white test score gap is identified from the relative performance of blacks and whites attending the same school, rather than across schools. To the extent that differential average school quality across races is the complete explanation for the widening racial test score gap, one would predict that the gap should not widen over time when comparing blacks and whites attending the same school. There are, of course, thorny issues of sample selection that potentially complicate the interpretation of these results: white students who elect to attend schools with black students might have had different test score trajectories than other white students, even if they had gone to all-white schools. Nonetheless, looking within schools provides a first attempt at testing this hypothesis.

The comparison of changes in the black-white test score gap over time including and excluding school-fixed effects is presented in table 6. All of the specifications in the table include the parsimonious set of covariates, although only the coefficient on the black-white gap is shown in the table. The first three columns reflect the full sample of students.

[21] Black students may attend these schools, but just not be in the classrooms sampled.

[22] Because elementary school students attend schools close to home, there is no way for us to distinguish between the impact of neighborhood and school quality in our data set. Note, however, that we are able to explain racial gaps upon entry to school without using controls for the neighborhood environment. For neighborhoods rather than schools to explain the racial divergence in test scores, the quality of the neighborhood would need to have a large impact on test scores after entry into school, but not before.

TABLE 7.—DIFFERENCES ACROSS RACES IN MEASURABLE SCHOOL INPUTS

School Input	Mean of School Input	Coefficient on Race in Predicting Level of School Input: Black	Hispanic	Asian	Other
Average class size	20.673 (3.875)	.591 (.340)	.699 (.271)	.799 (.349)	−.259 (.343)
Teacher has master's degree	0.280 (0.449)	.037 (.028)	.012 (.025)	−.001 (.032)	−.080 (.032)
Computer: student ratio	1.257 (2.050)	.003 (.156)	−.131 (.140)	.040 (.119)	.683 (.443)
Internet hookup: student ratio	0.344 (0.627)	−.048 (.037)	−.032 (.038)	.020 (.035)	.377 (.186)
Percentage of students in school with free lunch	29.83 (27.98)	19.32 (2.64)	8.17 (2.00)	3.27 (2.08)	6.81 (2.78)
Gang problems in school (1–3)	1.409 (0.585)	.261 (.058)	.338 (.044)	.128 (.044)	.336 (.069)
Problems with teacher turnover (1–5)	1.811 (0.943)	.263 (.083)	.227 (.064)	.062 (.078)	.132 (.092)
Litter around school (0–3)	0.741 (0.759)	.492 (.065)	.369 (.053)	.240 (.063)	.412 (.087)
People loitering around school (0–3)	0.524 (0.747)	.497 (.079)	.331 (.064)	.171 (.063)	.368 (.088)
Receives PTA funding	0.733 (0.442)	−.048 (.033)	−.050 (.026)	.000 (.029)	−.133 (.050)
Hall pass required	0.425 (0.494)	.194 (.037)	.100 (.034)	.010 (.041)	.059 (.046)

NOTES: The entries in the first column of the table are the means and standard deviations of the named school input. The entries in the remaining columns are estimated coefficients on race (with non-Hispanic whites as the omitted categories) from regressions of the named school inputs on the race dummies and other covariates included in the parsimonious set of controls. The method of estimation is weighted least squares using sample weights provided by ECLS. The reported standard errors have been corrected to take into account within-school correlation in the school-level measures.

The remaining columns restrict the sample to schools that have both black and white children in our sample. This set of students is relevant because only mixed-race schools provide useful variation to identify the racial test score gap when school fixed effects are included.

Column 3 of the table shows the baseline results reflecting the fact that blacks are losing ground in the full sample (−0.156 standard deviation relative to whites in math, −0.188 standard deviation in reading). When we eliminate students attending all-white schools from the sample, but otherwise estimate identical specifications, the results are not greatly affected (nor are they affected by eliminating students attending all-black schools). Blacks continue to lose substantial ground by the end of first grade. When school fixed effects are included in the regression (columns 7–9), the black-white test score gap is identified from differences between blacks and whites attending the same school. The estimate of ground lost by blacks shrinks to less than one-third of the magnitude in the full sample, and is not statistically different from 0 in these specifications.[23]

These findings are consistent with—but not definitive proof of—the argument that systematic differences in school quality for blacks and whites may explain the divergence in test scores. An alternative explanation is that whites who choose to attend schools with blacks are systematically worse than other whites. Note, however, that a comparison of columns 1 and 4 shows that in the fall of kindergarten black students actually fare somewhat worse relative to whites who attend schools with blacks then they do relative to the full sample of whites. This finding suggests that the whites who go to school with blacks (controlling for observables) actually achieve at a slightly higher level than do those who attend all-white schools, which is consistent with previous research. Moreover, comparing columns 4 and 7, we see that in kindergarten fall, blacks do even worse relative to whites attending the same school than they do relative to other whites. Thus, a simple selection story in which low-achieving whites are more likely to go to school with blacks is not consistent with the data. On the other hand, we cannot rule out a priori the possibility that whites who attend school with blacks are on lower academic trajectories, despite the fact that they initially score better on tests than other whites.

If blacks attend worse schools than whites on average, one might expect that this would be reflected in observable characteristics of the schools. Table 7 analyzes this issue. Each row of the table corresponds to a different measure of school quality. Column 1 presents means and standard deviations of each variable in the data, some of which are standard measures of school inputs (such as average class size and teacher education) and others of which are nontraditional (such as measures of gang problems and loitering). Unfortunately, the nontraditional measures are subjective responses by the school principal, administrator, or headmaster to questions of how serious problems such as gangs are at the school. Consequently, these measures are likely to be of poor quality. Columns 2–5 report the race coefficients from regressions that are parallel to those elsewhere in the paper, except that school inputs are the dependent variable rather than test scores. Thus, the entries in columns 2–5 reflect the extent to which children of other races attend higher or lower quality schools on each of the measures, controlling for our parsimonious set of covariates. On traditional measures of school quality such as class size, teacher's education, computers in class, and Internet connections, differences between blacks and whites are small. On the other hand, the percentage of students eligible for free lunch, the degree of gang problems in school, the amount of loitering in front of the school by nonstudents, and the amount of litter around the schools are much higher for blacks.

[23] This finding in some ways parallels Currie and Thomas's (1995) finding that early gains for students who attend Head Start tend to disappear due to low quality schools that these students later attend. Consistent with Currie and Thomas (2000), we do not find a positive effect of Head Start on student test scores even in kindergarten, once other factors are controlled for. This finding is also related to those of Krueger and Whitmore (2001) and Phillips, Crouse, and Ralph (1998b), who find that the black-white gap widens as a result of poorer quality schools.

There are important weaknesses in the argument that differential school quality explains the divergent trajectories of whites and blacks. First, the observable measures of school inputs included in table 7 explain only a small fraction of the variation in student outcomes. For instance, adding the school input measures to our basic student-level test score regressions only increases the R^2 of the regression by 0.05. Second, even after the school input measures are added to the test score regressions, the gap between blacks and whites continues to widen. Third, both Hispanics and Asians also experience worse schools than whites, but neither of those groups is losing ground. Because of these important weaknesses in the story—perhaps as a consequence of poor school quality measures in the data—the evidence linking school quality differences to the divergent trajectories of blacks can be characterized as no more than suggestive.

B. *Does the Importance of Parental and Environmental Inputs Grow As Children Age?*

Black children tend to grow up in environments less conducive to high educational attainment. If the importance of parental and environmental inputs grew as children aged, black students would be expected to lose ground relative to whites. The evidence in table 5, however, argues just the opposite. If that were true, than one would expect to observe the *raw* gaps widening between blacks and whites, but to the extent that our control variables adequately capture a child's environment, the residual gap after including all the covariates would remain constant. In fact, however, the residual gap increases more than the raw gap, contradicting this explanation.[24] Also, the magnitude of the coefficients on socioeconomic status, age at kindergarten entry, and mother's age at first birth are smaller in the first-grade test score regressions. That suggests that the relative importance of nonschool factors decreases over time, presumably because schools become a critical input into educational gains once children enter school.[25] Interestingly, the importance of school safety measures (gang problems, metal detectors, and the like) seem to become more important as children age.

[24] Indeed, from a theoretical perspective, one might expect that the opposite hypothesis would hold true: the importance of parental inputs declines with age. Prior to reaching school age, the relative share of educational inputs provided by parents is very large. Once school starts, much of the burden of educating is shifted to the schools. Our empirical evidence does not, however, provide much support for this conjecture either.

[25] An alternative explanation for the shrinking coefficient on the SES variable is that socioeconomic status varies over time, so that using the kindergarten value of the SES variable in the first-grade regression induces measurement error. That explanation cannot explain the declining coefficients on age at school entry and mother's age at birth. Moreover, for other variables that are time-varying, like number of books and WIC participation, the coefficients do not shrink in the first-grade regression.

TABLE 8.—DO BLACK STUDENTS SUFFER A GREATER SUMMER SETBACK WHEN SCHOOL IS NOT IN SESSION? ESTIMATES OF THE BLACK-WHITE TEST SCORE GAP FOR THE SUBSET OF THE SAMPLE TESTED IN FALL OF FIRST GRADE?

	Date Test Administered:			
Subject	Fall Kindergarten	Spring Kindergarten	Fall First Grade	Spring First Grade
Raw gaps:				
Math	−.601	**−.640**	**−.631**	−.696
	(.040)	(.044)	(.045)	(.048)
Reading	−.376	**−.421**	**−.390**	−.548
	(.042)	(.044)	(.043)	(.048)
With controls:				
Math	−.052	**−.097**	**−.134**	−.236
	(.040)	(.044)	(.045)	(.052)
Reading	.142	**.054**	**.071**	−.081
	(.043)	(.045)	(.044)	(.051)

Entries are coefficients on the variable black.
NOTES: Table entries are estimated black-white test score gaps at different points in time for the subset of the sample that has all four test scores. Only a small fraction of the sample was tested in fall of first grade. The total number of observations in the subsample is 5,223. The top panel of the table shows raw test score in boldface gaps; the bottom panel shows the residual test score gap, controlling for the parsimonious set of control variables. The observations in boldface represent the tests given shortly before and shortly after summer break. Standard errors are in parentheses.

C. *Do Black Children Suffer Worse Summer Setbacks When School Is Not in Session?*

Entwisle and Alexander (1992, 1994) and Heyns (1978) have argued that black students lose more ground over the summer than white students as a consequence of worse home and neighborhood environments, and they gain ground over the school year while in school. If this is the explanation for the falling performance of blacks, then public policies should be aimed not at schools, but rather, summer interventions. Our data provide a unique opportunity to test this hypothesis, because a subset of the sample is tested both in the spring of kindergarten and in the fall of first grade, shortly after students return to class, allowing us to isolate the relative summer setbacks for blacks and whites.

The results are reported in table 8. For the randomly chosen subset of the sample that is tested in the fall of first grade (approximately one-fourth of the students), we report at each point in time both the raw test score gap and the residual gap controlling for our parsimonious set of covariates. For the regression results, only the coefficient reflecting the black-white test score gap is shown in the table, and each entry in the table is from a separate regression. The test score gaps in the fall of kindergarten (column 1) and spring of first grade (column 4) for this subset of the sample are similar to those for the sample as a whole, suggesting that the subsample is indeed representative. Of greater interest is a comparison of the test scores in the spring of kindergarten versus the fall of first grade, for most of the intervening time was spent outside of school. On the raw scores, there is little difference before and after the summer break; to the extent there is any gap, it favors black students. With controls, black students lose slightly relative to whites over the summer in math (the gap rises from −0.097 to −0.134), but the null hypothesis of no change cannot be rejected. The

TABLE 9.—THE EVOLUTION OF THE PERFORMANCE GAP ON SUBJECTIVE TEACHER ASSESSMENTS FOR BLACKS, HISPANICS, AND ASIANS (ALL GAPS MEASURED RELATIVE TO WHITES)

Teacher's Subjective Assessment of Student Ability in:	Black-White Gap		Hispanic-White Gap		Asian-White Gap	
	Raw Data	Including All Controls	Raw Data	Including All Controls	Raw Data	Including All Controls
Math:						
Kindergarten fall	−.278 (.039)	−.103 (.038)	−.243 (.037)	−.097 (.034)	−.001 (.050)	.104 (.046)
First grade spring	−.463 (.062)	−.270 (.060)	−.242 (.049)	−.076 (.048)	.142 (.063)	.262 (.061)
Reading:						
Kindergarten fall	−.265 (.041)	−.066 (.040)	−.352 (.038)	−.177 (.034)	−.148 (.053)	−.041 (.049)
First grade spring	−.343 (.057)	−.146 (.054)	−.288 (.047)	−.101 (.045)	.063 (.062)	.190 (.058)

NOTES: The table entries are estimated racial gaps in subjective teacher assessments of student achievement provided by teachers in kindergarten fall and first-grade spring. The odd columns are raw gaps; the even columns are residual gaps after controlling for the parsimonious set of controls and teacher fixed effects. Teacher fixed effects are included in the regressions due to concerns over comparability across teachers in these assessments. Teacher assessments are normalized to have a mean of 0 and a standard deviations of 1 in the full, unweighted sample. The method of estimation is weighted least squares using sample weights provided by ECLS. Columns 2, 4, and 6 of each row are coefficients from one regression. The number of observations is 8,633 for math and 10,839 for reading.

point estimates for reading show slight gains by black students relative to whites over the summer.

Thus, the empirical results lend little support to the hypothesis that differential summer setbacks explain the lost ground of black students in our sample. We do observe blacks losing ground during the school year in both subjects in both years, in direct conflict with Entwisle and Alexander (1992).

D. *Are the Results Simply an Artifact of Standardized Testing?*

Given the potential difficulties of evaluating student achievement using standardized tests in children so young, one possibility is that our results are simply an artifact of standardized testing. To assess this hypothesis, we examined the relative performance of children of different races on subjective teacher evaluations. The teacher assessments were normalized to mean 0 and standard deviation 1. The regressions are identical to the specifications using test scores, with one important difference: Because of concerns about heterogeneity across teachers in the way they may rate students on these subjective evaluations, we include teacher fixed effects in all of the regressions using these measures. Thus, the estimates are based on a student's evaluation relative to other students in the same classroom.

The results for teacher assessments are presented in table 9. The odd columns contain the raw gaps across races on the assessments; the even columns control for our parsimonious set of regressors. As before, coefficients are in standard deviation units. In kindergarten fall, black students are judged by teachers to be 0.26 to 0.28 standard deviations behind whites on both math and reading. These gaps are smaller than those observed on the standardized tests. Controlling for other characteristics, the black-white gap shrinks to 0.103 standard deviation in math and 0.066 standard deviation in reading. As was the case with test scores, the gap between whites and blacks widens substantially by the end of first grade. Thus, the patterns in the test scores are replicated in teacher assessments. This is true not only for blacks, but also for Hispanics and Asians. The most notable divergence in results between test scores and teacher assessments is that teachers initially rate Asians as performing no better than whites, but increase their evaluation of Asians over time. In the test score data, the opposite pattern emerges.

E. *Can Teacher Bias or Differential Socialization Explain Black Students Losing Ground?*

If, as some have argued, white teachers have lower expectations for black children or otherwise discriminate against them in the classroom (Baron, Tom, & Cooper, 1985; Dusek, 1975; Ferguson, 1998; Lightfoot, 1978), then one would predict that black students with white teachers should lose more ground than black students with black teachers. Table 10 tests this theory. Columns 1 and 2 correspond to the subsample of students neither of whose teachers in kindergarten or first grade are black. Columns 3 and 4 reflect students who have at least one black teacher. Note that less than 5% of white students have a black teacher by the end of first grade, compared to over 50% of black students. The top rows of the table report test scores without teacher fixed effects, the middle panel reports test scores including teacher fixed effects, and the bottom panel reports subjective teacher assessments (with teacher fixed effects). The results are generally similar across the three sets of analysis. Black children who have at least one black teacher start out somewhat worse relative to their white peers on math, and slightly better on reading, than black students who have no black teachers. By the end of first grade, however, the black-white test score gap is greater across the board for students who have at least one black teacher (that is, the coefficients in column 4 are always more negative than those in column 2). This finding is exactly the opposite of what one would predict from a discrimination story.

Although we do not show the analysis in tabular form, we have also explored whether differences in socialization and behavior might explain the results. The data set includes a number of variables that might be correlated with cultural differences in socialization. Parents and teachers were asked to rate the child on various social skills, including self-control, approaches to learning, interpersonal skills, and exhibiting problem behaviors. The inclusion of these variables in our regressions, however, had virtually no impact

TABLE 10.—DIFFERENCES IN THE EVOLUTION OF THE TEST SCORE GAP BY TEACHER'S RACE

Subject	Neither Kindergarten nor First Grade Teacher is Black		Kindergarten and/or First Grade Teacher is Black	
	Fall Kindergarten	Spring First Grade	Fall Kindergarten	Spring First Grade
Full sample (without fixed effects):				
Math	−.134 (.031)	−.251 (.036)	−.149 (.079)	−.378 (.085)
Reading	.081 (.038)	−.072 (.037)	.163 (.073)	−.172 (.085)
Full sample (with fixed effects):				
Math	−.162 (.041)	−.237 (.045)	−.218 (.098)	−.367 (.113)
Reading	−.015 (.048)	−.053 (.048)	.115 (.109)	−.182 (.116)
Teacher assessments (with fixed effects):				
Math teacher assessment	−.125 (.053)	−.206 (.051)	−.154 (.102)	−.299 (.136)
Reading teacher assessment	−.135 (.048)	−.109 (.052)	−.183 (.112)	−.248 (.142)
Number of white observations	6,885		371	
Number of black observations	863		1,170	

NOTES: The entries are estimated black-white test score gaps in the top two panels and subjective teacher assessments in the third panel. Students in the data set are assigned to two mutually exclusive groups in this table: students who do and do not have any black teachers in kindergarten or first grade. The number of observations falling into each of those categories, by race of the student, is reported at the bottom of the table. The top panel of the table reports results from regressions including the parsimonious set of controls. The second panel adds teacher fixed effects. The bottom panel is teacher assessments, controlling for teacher fixed effects.

on the measured black-white test score gap or its trajectory over time.[26]

F. *Does the Material Tested Change As Children Age in a Manner That Lowers the Relative Performance of Black Students?*

In personal correspondence, ECLS reports that the fraction of the exam devoted to each set of skills remains constant as children age. In the fall of kindergarten, few children are expected to correctly answer questions involving multiplication and division, and by the spring of first grade, few children are expected to miss questions involving counting. Nonetheless, the same mix of questions is asked.

Still, it is potentially interesting to compare the relative performance of black and white children on the different types of questions over time. Table 11 reports the unadjusted means, by race, of children in fall kindergarten and spring first grade on questions assessing specific sets of skills. These numbers are quite illuminating. Black and white children enter kindergarten with similar scores in counting, numbers, and shapes and show nearly complete mastery of these skills by the end of first grade. In contrast, blacks score substantially below whites on multiplying and dividing when they enter kindergarten, and these differences are exaggerated. Table 12 reports the black-white gap on questions assessing specific sets of skills at both the beginning and the end of the sample. The estimates in the table are based on specifications including our parsimonious set of regressors, and the magnitude of the estimates is once again expressed in standard deviation units. In math, the lost ground on the part of blacks is attributable almost solely to poor performance in addition, subtraction, multiplication, and division. Black students perform almost as well as whites in the fall of kindergarten on these tasks, but lag whites by over two-tenths of a standard deviation by the spring of first grade. To the extent that skill in manipulating numbers is likely to be a more important input into understanding higher-order math than are simple tasks such as counting or evaluating relative size, these results may foreshadow continued losses for blacks relative to whites in math as they age. Blacks lose roughly equal ground on all aspects of reading proficiency.

[26] Although we do not directly test other social-theoretic explanations such as stereotype threat and acting white, it would seem unlikely that they would affect children at such an early age.

TABLE 11.—UNADJUSTED MEANS ON QUESTIONS ASSESSING SPECIFIC SETS OF SKILLS

Skill Tested	Fall Kindergarten		Spring First Grade	
	Black	White	Black	White
Math:				
Count, number, shapes	.885 (.228)	.956 (.139)	.997 (.037)	.999 (.013)
Relative size	.415 (.352)	.647 (.350)	.967 (.119)	.990 (.062)
Ordinality, sequence	.095 (.207)	.269 (.342)	.882 (.256)	.963 (.143)
Add/subtract	.010 (.059)	.053 (.151)	.572 (.357)	.795 (.289)
Multiply/divide	.001 (.021)	.005 (.049)	.097 (.206)	.324 (.358)
Reading:				
Letter recognition	.568 (.434)	.719 (.395)	.990 (.082)	.997 (.044)
Beginning sounds	.200 (.305)	.350 (.375)	.934 (.173)	.977 (.099)
Ending sounds	.101 (.211)	.204 (.296)	.865 (.241)	.945 (.145)
Sight words	.011 (.087)	.031 (.145)	.671 (.411)	.842 (.310)
Words in context	.004 (.048)	.012 (.092)	.311 (.379)	.505 (.413)

NOTES: Entries are unadjusted mean scores on specific areas of questions in kindergarten fall and first-grade spring. They are proficient probability scores, which are constructed using IRT scores and provide the probability of mastery of a specific set of skills.

TABLE 12.—BLACK-WHITE PERFORMANCE GAPS ON QUESTIONS ASSESSING SPECIFIC SETS OF SKILLS

	Black-White Gap	
Skill Tested	Fall Kindergarten	Spring First Grade
Math:		
Count, number, shapes	.042 (.030)	−.020 (.045)
Relative size	−.108 (.028)	−.035 (.038)
Ordinality, sequence	−.108 (.023)	−.067 (.035)
Add/subtract	−.056 (.018)	−.232 (.031)
Multiply/divide	.005 (.011)	−.262 (.023)
Reading:		
Letter recognition	.114 (.030)	.033 (.047)
Beginning sounds	.099 (.027)	−.038 (.036)
Ending sounds	.107 (.025)	−.068 (.034)
Sight words	.072 (.025)	−.090 (.033)
Words in context	.057 (.022)	−.061 (.029)

NOTES: Entries in the table are residual black-white test score gaps on specific types of questions in kindergarten fall and first-grade spring. On each category of question, student scores have been normalized to have a mean of 0 and a standard deviation of 1 in the full, unweighted sample. The parsimonious set of controls is included in all regressions. The method of estimation is weighted least squares using sample weights provided by ECLS. The number of observations on reading is 13,290, and on math is 12,601.

VIII. Conclusion

Previous efforts to explain the black-white test score gap have generally fallen short—a substantial residual remained for black students, even after controlling for a full set of available covariates. Using a new data set, we demonstrate that among entering kindergartners, the black-white gap in test scores can be essentially eliminated by controlling for just a small number of observable characteristics of the children and their environment. Once students enter school, the gap between white and black children grows, even conditional on observable factors. We test a number of possible explanations for why blacks lose ground. We speculate that blacks are losing ground relative to whites because they attend lower quality schools, though we recognize that we have not provided definitive proof. This is the only hypothesis which receives any empirical support. To test this hypothesis convincingly, we need more detailed data on schools, neighborhoods, and the general environment kids grow up in.

Compared to previous studies, our results provide reason for optimism. Research on earlier cohorts of children found much greater black-white test score gaps, both in the raw scores and controlling for observables. When we attempt to mimic the nonrandom sample frames in earlier research (for example only looking at low-birth-weight babies, as in IHDP), we continue to find much smaller gaps in our sample. One plausible explanation for the differences between the current sample and cohorts attending kindergarten 10–30 years ago is that the current cohort of blacks has made real gains relative to whites. Recent cohorts show smaller black-white gaps in the raw data, across multiple data sets, which gives reason for optimism.

REFERENCES

Armor, G., "Why is Black Educational Achievement Rising?" *Public Interest* (September 1992), 65–80.

Baron, R., D. Tom, and H. Cooper, "Social Class, Race, and Teacher Expectations," in Jerome Dusek (Ed.), *Teacher Expectancies* (Hillsdale, NJ: Erlbaum, 1985).

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

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

Brooks-Gunn, J., and G. J. Duncan (Eds.), *The Consequences of Growing Up Poor* (New York: Russell Sage, 1997).

Brooks-Gunn, J., and G. J. Duncan, "Family Poverty, Welfare Reform and Child Development," *Child Development* 71:1 (2000), 188–196.

Brooks-Gunn, J., G. J. Duncan, and P. Klebanov, "Economic Deprivation and Early-Childhood Development," *Child Development* 65:2 (1994), 296–318.

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

Brooks-Gunn, J., G. J. Duncan, P. Klebanov, and N. Sealand, "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology* 99:2 (1993a), 353–395.

Brooks-Gunn, J., P. K. Klebanov, and G. K. Duncan, "Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics," *Child Development* 67:2 (1995), 396–408.

Brooks-Gunn, J., P. K. Klebanov, F. Liaw, and D. Spiker, "Enhancing the Development of Low Birth-Weight, Premature Infants: Changes in Cognition and Behavior over the First Three Years," *Child Development* 63 (1993b), 736–753.

Carneiro, P., and J. Heckman, "Human Capital Policy," University of Chicago working paper (2002).

Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. D. Weinfeld, and P. York, *Equality of Educational Opportunity* (Washington, DC: U.S. Government Printing Office, 1966).

Coley, R. J., "An Uneven Start: Indicators of Inequality in School Readiness," Educational Testing Service report (March 2002).

Cook, M., and W. Evans, "Families or Schools? Explaining the Convergence in White and Black Academic Performance," *Journal of Labor Economics* 18:4 (2000), 729–754.

Cook, P., and J. Ludwig, "The Burden of Acting White: Do Black Adolescents Disparage Academic Achievement?" (pp. 375–400), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998).

Crane, J., "Race and Children's Cognitive Test Scores: Empirical Evidence That Environment Explains the Entire Gap," University of Illinois at Chicago (1994).

Currie, J., and D. Thomas, "School Quality and the Longer Term Effects of Head Start," *Journal of Human Resources* 35:4 (1995), 755–774.

Delpit, L., *Other People's Children: Cultural Conflict in the Classroom* (New York: The New Press, 1995).

Devlin, B., D. Resnick, and K. Roeder, *Intelligence, Genes, and Success: Scientists Respond to the Bell Curve* (Copernicus Books, 1998).

Dusek, J. B. "Do Teachers Bias Children's Learning?" *Review of Educational Research* 45 (1975), 661–684.

Entwisle, D. R., and K. Alexander, "Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School," *American Sociological Review* 57 (1992), 72–84.

——— "Winter Setback: The Racial Composition of Schools and Learning to Read," *American Sociological Review* 59 (1994), 446–460.

Ferguson, R. F., "Teachers' Perceptions and Expectations and the Black-White Test Score Gap" (pp. 273–317), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998).

Fordham, S., and J. Ogbu, "Black Students' School Successes: Coping with the Burden of Acting White," *The Urban Review* 18:3 (1986), 176–206.

Fraser, S., *The Bell Curve Wars: Race, Intelligence, and the Future of America* (New York: Basic Books, 1995).

Fryer, R., "An Economic Approach to Cultural Capital," University of Chicago working paper (2002).

Goldberg, A., and C. Manski, "Review Article: The Bell Curve," *Journal of Economic Literature* 33:2 (1995), 762–776.

Grissmer, D., A. Flanagan, and S. Williamson, "Why Did the Black-White Score Gap Narrow in the 1970's and 1980's?" (pp. 182–228), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998).

Grogger, J., and D. Neal, "Further Evidence on the Benefits of Catholic Secondary Schooling," *Brookings-Wharton Papers on Urban Affairs* 1 (2000), 151–193.

Heckman, J. J., "Lessons from the Bell Curve," *Journal of Political Economy* 103:5 (1995), 1091–1120.

Hedges, L., and A. Nowell, "Black-White Test Score Convergence since 1965" (pp. 149–181), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998).

Hernstein, R. J., and C. Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (The Free Press, 1994).

Heyns, B., *Summer Learning and the Effects of Schooling* (New York: Academic Press, 1978).

Humphreys, L., "Trends in Levels of Academic Achievement of Blacks and Other Minorities," *Intelligence* 12 (1988), 231–260.

Jencks, C., and M. Phillips, "America's Next Achievement Test: Closing the Black-White Test Score Gap," *American Prospect* 40 (September–October 1998), 44–53.

Jensen, A., "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review* 39 (1969), 1–123.

——— *Educability and Group Differences* (New York: Free Press, 1973).

——— *The G Factor: The Science of Mental Ability* (Greenwood Publishing Group, 1998).

Jones, L. V., N. Burton, and E. Davenport, *Mathematics Achievement Levels of Black and White Youth* (Chapel Hill: University of North Carolina, L. L. Thurstone Psychometric Laboratory, 1982).

Kaufman, A., and N. Kaufman, *K-ABC: Kaufman Assessment Battery for Children* (Circle Pines, MN: American Guidance Services, 1983).

Kincheloe, J., S. Steinberg, and A. Gresson, *Measured Lies: The Bell Curve Re-examined* (St. Martins Press, 1997).

Krohn, E., and R. Lamp, "Current Validity of the Stanford-Binet Fourth Edition and K-ABC for Head Start Children," *Journal of Psychology* 27 (1989), 59–67.

Krueger, A., and D. Whitmore, "Would Smaller Classes Help Close the Black White Achievement Gap?" Industrial Relations Section, Princeton University, working paper no. 451 (2001).

Li, K., and D. J. Poirier, "The Roles of Birth Inputs and Outputs in Predicting Health, Behavior, and Test Scores at Age Five or Six," University of British Columbia working paper (2001).

Lightfoot, S., *Worlds Apart: Relationships between Families and Schools* (New York: Basic Books, 1978).

Mayer, S. E., *What Money Can't Buy: Family Income and Children's Life Chances* (Harvard University Press, 1997).

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

National Center of Educational Statistics, *User's Guide to the Kindergarten–First Grade Public Use Data File,* NCES-2002-149 (2002).

Neal, D., "The Effects of Catholic Secondary Schooling on Educational Achievement," *Journal of Labor Economics* 15 (1997), 98–123.

Neal, D., and W. R. Johnson, "The Role of Pre-market Factors in Black-White Wage Differences," *Journal of Political Economy* 104 (1996), 869–895.

O'Neill, J., "The Role of Human Capital in Earnings Differences between Black and White Men," *Journal of Economic Perspectives* 4:4 (1990), 25–46.

Phillips, M., "Understanding Ethnic Differences in Academic Achievement: Empirical Lessons from National Data" (pp. 103–132), in D. Grissmer and M. Ross (Eds.), *Analytic Issues in the Assessment of Student Achievement* (Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2000).

Phillips, M., J. Brooks-Gunn, G. J. Duncan, and P. K. Klebanov, "Family Background, Parenting Practices, and the Black-White Test Score Gap" (pp. 103–145), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998a).

Phillips, M., J. Crouse, and J. Ralph, Does the Black-White Test Score Gap Widen after Children Enter School?" (pp. 229–272), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998b).

Rodgers, W., and W. Spriggs, "What Does AFQT Really Measure: Race, Wages, Schooling and the AFQT Score," *The Review of Black Political Economy* 24:4 (1996), 13–46.

Scarr, S., *Race, Social Class and Individual Differences in I.Q.* (Lawrence Erlbaum Associates, 1981).

Steele, C., and J. Aronson, "Stereotype Threat and the Test Performance of Academically Successful African Americans" (pp. 401–430), in C. Jencks and M. Phillips (Eds.), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institute, 1998).

DATA APPENDIX

The Early Childhood Longitudinal Study kindergarten cohort (ECLS-K) is a nationally representative sample of 21,260 children entering kindergarten in 1998. Thus far, information on these children has been gathered at four separate points in time. The full sample was interviewed in the fall and spring of kindergarten and spring of first grade. All of our regressions and summary statistics are weighted, unless otherwise noted, and we include dummies for missing data. We describe below how we combined and recoded some of the ECLS variables used in our analysis.

1. Socioeconomic Composite Measure

The socioeconomic scale variable (SES) was computed by ECLS at the household level for the set of parents who completed the parent interview in fall kindergarten or spring kindergarten. The SES variable reflects the socioeconomic status of the household at the time of data collection for spring kindergarten. The components used for the creation of SES were: father or male guardian's education, mother or female guardian's education, father or male guardian's occupation, mother or female guardian's occupation, and household income.

2. Number of Children's Books

Parents or guardians were asked, "How many books does your child have in your home now, including library books?" Answers ranged from 0 to 200.

3. Child's Age

We used the composite variable *child's age at assessment* provided by ECLS. The child's age was calculated by determining the number of days between the child assessment date and the child's date of birth. The number was then divided by 30 to calculate the age in months.

4. Birth Weight

Parents were asked how much their child weighed when they were born. We multiplied the number of pounds by 16 and added it to the ounces to calculate birth weight in ounces.

5. Mother's Age at First Birth

Mothers were asked how old they were at the birth of their first child.

6. Average Class Size

We computed each child's average class size over their kindergarten year by adding their class size in the fall and spring and dividing by two.

7. Teacher Has Master's Degree

We coded a dummy variable equal to 1 if the child's teacher has a master's degree or above.

8. Computer: Student Ratio

The number of computers in each school and the total enrollment of each kindergarten program are provided by the ECLS, based on a survey given to each school. We divided the number of computers in each school by the total enrollment in kindergarten to produce this ratio.

9. Internet Hookup: Student Ratio

This was constructed similar to the computer: student ratio, except the numerator consists of the number of Internet and LAN connections in the school.

10. Percentage of Students in Child's School Available for Free Lunch

Schools provided the percentage of students in their school who were eligible for free lunch.

11. Gang Problems

Schools were asked, "How much of a problem are gangs in the neighborhood where the school is located?" We coded this variable so that 1 implies "no problem," 2 implies "something of a problem," and 3 implies "big problem."

12. Teacher Turnover

Schools were asked how much they agreed with the statement "Teacher turnover is a problem in this school." Answers range from 0 to 5, 0 indicating they strongly disagree and 5 indicating they strongly agree.

13. Litter around School

The ECLS interviewer was asked to report the amount of litter around each school. The variable ranges from 0 to 3, where 0 indicates no litter and 3 indicates "a lot."

14. People Loitering around School

The ECLS interviewer was asked to report the amount of loitering by nonstudents around the school. The variable ranges from 0 to 3, where 0 indicates none, and 3 indicates "a lot."

15. PTA Funding

Schools reported whether or not they receive supplemental funding from their PTA. We recoded this variable so that 1 implies yes and 0 implies no.

16. Hall Pass Required

Schools were asked, "Are hall passes required to ensure the safety of the children in your school?" This variable is coded 1 if yes and 0 if no.

TABLE A1.—SUMMARY STATISTICS BY RACE: STUDENT CHARACTERISTICS

Variable	Black	White
Baseline child characteristics:		
Female	0.495 (0.536)	0.481 (0.594)
Age, fall of kindergarten	66.877 (4.776)	67.402 (4.752)
Geography:		
Northeast region	0.134 (0.341)	0.214 (0.396)
Midwest region	0.170 (0.390)	0.285 (0.495)
South region	0.613 (0.536)	0.344 (0.495)
West region	0.083 (0.292)	0.156 (0.396)
Urban	0.533 (0.536)	0.418 (0.495)
Suburban	0.359 (0.536)	0.323 (0.495)
Rural	0.108 (0.341)	0.260 (0.495)
Family composition:		
Two parents, both biological	0.317 (0.536)	0.738 (0.495)
Two parents, one biological	0.083 (0.292)	0.091 (0.297)
Single parent	0.502 (0.536)	0.147 (0.396)
Adopted	0.019 (0.146)	0.011 (0.099)
In custody of guardian	0.079 (0.292)	0.013 (0.099)
Number of siblings in home	1.595 (1.462)	1.369 (1.089)
Mother's age at first birth	20.548 (5.139)	24.579 (5.916)
Mother age at this child's birth	26.587 (9.169)	28.046 (6.760)
Parental education and income:		
Mother's level of education	12.682 (2.130)	13.861 (2.454)
Father's level of education	12.964 (2.246)	13.993 (2.798)
SES measure	−0.359 (0.780)	0.202 (0.792)
WIC participation (1 = yes)	0.772 (0.481)	0.332 (0.493)
Food stamp participation	0.451 (0.533)	0.103 (0.395)
Mother's occupational status	40.642 (10.512)	44.930 (12.648)
Father's occupational status	39.698 (9.076)	44.111 (11.949)
Child's early home environment:		
Child's birth weight (in ounces)	110.315 (25.373)	120.256 (22.942)
Working mother	0.817 (0.420)	0.770 (0.487)
No Nonparental care	0.135 (0.384)	0.171 (0.394)
Relative care	0.179 (0.432)	0.110 (0.296)
Nonrelative care	0.041 (0.240)	0.126 (0.394)
Center-based program	0.337 (0.528)	0.493 (0.591)
Head Start participation	0.227 (0.480)	0.055 (0.296)
Varied		
English spoken at home	0.991 (0.097)	0.986 (0.099)
Number of books in home	39.014 (41.986)	93.121 (64.792)
Neighborhood characteristics (1 = "no problem," 3 = "big problem"):		
How big a problem is personal safety?	1.515 (0.013)	1.217 (0.005)
How big a problem are drugs?	1.30 (0.013)	1.069 (0.003)
How big a problem is burglary?	1.164 (0.010)	1.103 (0.004)
How big a problem is violence?	1.093 (0.008)	1.017 (0.002)
How big a problem are vacant houses?	1.140 (0.009)	1.038 (0.003)
Parental involvement (1 = "not at all," 4 = "every day"):		
How often does parent read books to child?	3.009 (0.924)	3.352 (0.791)
How often does parent tell stories to child?	2.644 (1.021)	2.757 (0.989)
In the past month, has parent taken child to library? (1 = "yes")	0.486 (0.535)	0.559 (0.594)
In the past month, has parent taken child to museum?	0.285 (0.487)	0.329 (0.495)
Has parent attended a school open house this school year?	0.593 (0.536)	0.792 (0.494)
Has parent attended a PTA meeting this school year?	0.357 (0.536)	0.334 (0.495)
Has parent volunteered at school this school year?	0.287 (0.487)	0.562 (0.495)
Kindergarten program:		
Morning program—fall	0.128 (0.390)	0.295 (0.495)
Afternoon program—fall	0.094 (0.341)	0.192 (0.396)
All day program—fall	0.777 (0.439)	0.513 (0.594)
Morning program—spring	0.128 (0.379)	0.296 (0.682)
Afternoon program—spring	0.089 (0.332)	0.192 (0.390)
All day program—spring	0.783 (0.474)	0.513 (0.585)
Frequency of missing values for key variables:		
Mother's age	0.045 (0.244)	0.008 (0.099)
Mother's education	0.011 (0.097)	0.017 (0.099)
Father's education	0.509 (0.536)	0.114 (0.396)
Mother's occupation	0.296 (0.487)	0.328 (0.495)
Father's occupation	0.597 (0.536)	0.170 (0.396)

TABLE A2.—FULL REGRESSION RESULTS OF BASELINE SPECIFICATIONS: FALL KINDERGARTEN

Variable	Math Full Sample	Reading Full Sample
Race:		
Black	−0.102 (0.026)	0.093 (0.030)
Hispanic	−0.171 (0.028)	−0.076 (0.029)
Asian	0.274 (0.050)	0.375 (0.060)
Other	−0.113 (0.035)	−0.014 (0.039)
Geographic controls:		
Northeast	0.070 (0.025)	0.034 (0.027)
Midwest	0.021 (0.022)	−0.047 (0.023)
West	0.060 (0.025)	0.024 (0.028)
Rural	−0.166 (0.022)	−0.186 (0.023)
Suburban	−0.071 (0.018)	−0.094 (0.019)
Baseline child characteristics:		
Female	0.000 (0.015)	0.153 (0.016)
Age (in months)	−2.680 (0.542)	−2.409 (0.483)
Age^2	0.041 (0.008)	0.037 (0.007)
Age^3 × 1,000	−0.209 (0.039)	−0.181 (0.033)
Home environment:		
Sibling	−0.058 (0.024)	−0.110 (0.027)
Sibling^2	0.004 (0.009)	0.006 (0.010)
Sibling^3	−0.000 (0.001)	0.000 (0.001)
Biological mother and other father	−0.031 (0.031)	−0.050 (0.031)
Other mother and biological father	−0.101 (0.089)	−0.118 (0.080)
Biological mother only	−0.033 (0.048)	−0.089 (0.058)
Biological father only	0.133 (0.152)	0.010 (0.183)
Adopted	−0.226 (0.065)	−0.056 (0.079)
Guardian	−0.061 (0.088)	−0.023 (0.104)
Missing mother's age at 1st birth	−0.086 (0.059)	−0.100 (0.071)
Mother in her teens at 1st birth	0.029 (0.021)	−0.069 (0.022)
Mother older than 30 at 1st birth	0.111 (0.028)	0.155 (0.030)
Missing mother's age at child's birth	−0.218 (0.104)	−0.064 (0.144)
Mother in her teens at child's birth	−0.041 (0.024)	−0.022 (0.025)
Mother older than 30 at child's birth	−0.016 (0.021)	−0.049 (0.021)
Mother high school graduate	0.048 (0.025)	0.045 (0.026)
Mother attended vocational school	0.129 (0.041)	0.094 (0.043)
Mother has some college	0.100 (0.032)	0.086 (0.033)
Mother has bachelor's degree	0.230 (0.041)	0.178 (0.042)
Mother has graduate degree	0.279 (0.052)	0.241 (0.056)
Missing mother's education	0.105 (0.119)	0.032 (0.135)
Father high school graduate	0.064 (0.027)	0.043 (0.030)
Father attended vocational school	0.093 (0.045)	0.134 (0.053)
Father has some college	0.117 (0.033)	0.099 (0.036)
Father has bachelor's degree	0.192 (0.039)	0.169 (0.041)
Father has graduate degree	0.338 (0.051)	0.303 (0.054)
Missing father's education	0.089 (0.055)	0.146 (0.063)
Socioeconomic status	0.072 (0.024)	0.092 (0.023)
WIC	−0.120 (0.020)	−0.104 (0.021)
Missing WIC	−0.121 (0.085)	−0.244 (0.102)
Food stamp	−0.075 (0.022)	−0.075 (0.024)
Missing food stamp	−0.044 (0.109)	−0.000 (0.128)
Relative preschool care	0.020 (0.027)	−0.004 (0.029)
Nonrelative preschool care	0.081 (0.032)	0.036 (0.034)
Center-based preschool care	0.152 (0.023)	0.150 (0.025)
Head start	0.016 (0.030)	−0.043 (0.030)
Varied preschool care	0.112 (0.040)	0.072 (0.041)
Missing preschool care	0.196 (0.075)	0.064 (0.072)
Mother did not work between birth and kindergarten	0.010 (0.021)	0.004 (0.022)
Missing whether mother worked between birth and kindergarten	0.018 (0.084)	−0.023 (0.093)
Birth weight × 10	0.030 (0.004)	0.022 (0.004)
Missing birth weight	0.279 (0.050)	0.179 (0.053)
Neighborhood safety	−0.115 (0.051)	0.024 (0.057)
Neighborhood drug use	−0.060 (0.073)	−0.060 (0.082)
Neighborhood burglary	0.040 (0.080)	0.066 (0.090)
Neighborhood violence	0.086 (0.112)	−0.106 (0.136)
Neighborhood vacancies	−0.046 (0.094)	−0.151 (0.103)
Mother's occupation	0.107 (0.056)	0.070 (0.061)
(Mother's occupation)^2	−0.002 (0.001)	−0.001 (0.001)
(Mother's occupation)^3 × 1,000	0.012 (0.007)	0.008 (0.008)
Missing mother's occupation	1.872 (0.918)	1.265 (0.990)
Father's occupation	−0.017 (0.046)	0.030 (0.048)
(Father's occupation)^2	0.000 (0.001)	−0.001 (0.001)
(Father's occupation)^3 × 1,000	−0.003 (0.006)	0.003 (0.006)
Missing father's occupation	−0.209 (0.753)	0.458 (0.791)
English spoken in home	0.117 (0.031)	0.091 (0.045)
Missing English	−0.296 (0.187)	−0.146 (0.203)
Read book to child once or twice a week	0.071 (0.071)	−0.070 (0.112)
Read book to child 3–6 times a week	0.091 (0.072)	−0.061 (0.114)
Read book to child every day	0.135 (0.073)	0.064 (0.114)
Missing read book	0.562 (0.226)	0.479 (0.202)
Tell stories once or twice a week	0.074 (0.030)	0.064 (0.032)
Tell stories 3–6 times a week	0.096 (0.032)	0.056 (0.033)
Tell stories every day	0.026 (0.033)	0.014 (0.035)
Missing tell story	−0.293 (0.171)	−0.273 (0.157)
Number of books in home	0.005 (0.001)	0.004 (0.001)
(Number of books in home)^2 ×1,000	−0.027 (0.016)	−0.017 (0.018)
(Number of books in home)^3 ×1,000	0.000 (0.000)	0.000 (0.000)
Missing number of books	0.251 (0.087)	0.131 (0.088)
Do not take child to library	−0.047 (0.016)	−0.041 (0.017)
Missing library	−0.389 (0.185)	−0.158 (0.260)
Do not take child to museum	0.018 (0.017)	0.015 (0.018)
Missing museum	−0.181 (0.226)	−0.395 (0.337)
Do not go to open house	−0.030 (0.018)	−0.000 (0.020)
Missing open house	−0.233 (0.135)	−0.136 (0.142)
Do not go to PTA	−0.036 (0.017)	−0.038 (0.019)
Missing PTA	0.564 (0.243)	0.542 (0.273)
Do not volunteer in school	−0.061 (0.017)	−0.054 (0.018)
Missing volunteer	−0.217 (0.190)	−0.376 (0.219)
Never spank child	0.037 (0.038)	0.043 (0.042)
Spanked child 0 times last week	0.053 (0.035)	0.038 (0.038)
Spanked child *x* times last week	−0.027 (0.026)	−0.014 (0.028)
Spanked child *x* times last week (squared)	0.005 (0.003)	0.003 (0.004)
(Spanked child *x* times last week)^3 ×10	−0.002 (0.001)	−0.001 (0.001)
Missing spank	−0.068 (0.092)	0.005 (0.113)
Morning kindergarten program—fall	−0.375 (0.116)	−0.386 (0.108)
Afternoon kindergarten program—fall	−0.443 (0.097)	−0.376 (0.116)
Morning kindergarten program—spring	0.292 (0.116)	0.280 (0.108)
Afternoon kindergarten program—spring	0.376 (0.097)	0.297 (0.116)

NOTES: See notes to table 2. The two columns in this table report the full results of specifications 5 and 10 reported in table 2.