The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps

Author(s): Petra E. Todd and Kenneth I. Wolpin

Source: *Journal of Human Capital*, Vol. 1, No. 1 (Winter 2007), pp. 91-136

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/10.1086/526401

The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps

Petra E. Todd and Kenneth I. Wolpin
University of Pennsylvania

This paper studies the determinants of children's scores on tests of cognitive achievement in math and reading. Using rich longitudinal data on test scores, home environments, and schools, we implement alternative specifications for the cognitive achievement production function that allow achievement to depend on the entire history of lagged home and school inputs as well as on parents' ability and unobserved endowments. We use cross-validation methods to select among competing specifications and find support for a variant of a value-added model of the production function augmented to include information on lagged inputs. Using this specification, we study the sources of test score gaps between black, white, and Hispanic children. The estimated model captures key patterns in the data, such as the widening of minority-white test score gaps with age and differences in the gap pattern between Hispanics and blacks. We find that differences in mother's "ability," as measured by AFQT, account for about half of the test score gap. Home inputs also account for a significant proportion. Equalizing home inputs at the average levels of white children would close the black-white and the Hispanic-white test score gaps in math and reading by about 10–20 percent.

I. Introduction

Accumulated evidence finds that early test scores are predictive of future labor market success. For example, scores on cognitive achievement tests taken by adolescents are predictive of their future educational attainment, labor force participation, earnings, and even incarceration (see, e.g., Leibowitz 1974; Murnane, Willett, and Levy 1995; Neal and Johnson 1996; Keane and Wolpin 1997; Cameron and Heckman 1998; Hanushek and Rivkin 2006). Scores on tests taken as early as age 7 have been shown to be correlated with education, earnings, and occupational

We thank Isaac Ehrlich, Rick Hanushek, James Heckman, Michael Keane, Richard Murnane, Derek Neal, and Robert Tamura for helpful comments. This paper was presented at the 2006 inaugural conference of the *Journal of Human Capital* in Buffalo, at Columbia University, at the Institute for Research on Poverty conference in Madison, WI, and at Washington University in St. Louis. An earlier version was presented at seminars at Harvard University, Johns Hopkins University, Harris School of Public Policy (Chicago), University of California at Los Angeles and San Diego, University of Toulouse, University of Bergen, and a conference at the Minneapolis Federal Reserve Bank.

[*Journal of Human Capital*, 2007, vol. 1, no. 1]

choices.[1] This evidence along with the observed racial disparities in labor market outcomes has led many researchers to assign a large role to "premarket factors" in explaining the disparities. Premarket factors are broadly interpreted to represent family and school influences on skills and abilities valued in the labor market.

Premarket factors are also thought to be an important part of the explanation for racial disparities in test score performance (Neal and Johnson 1996). Although it is conceivable that test score gaps could arise from differential investment in children based on expectations about future labor market returns (a postmarket rather than a premarket factor), Carniero, Heckman, and Masterov (2002) argue that this is unlikely. They document that test score gaps between white and black children already emerge by the age of school entry and tend to widen with age.[2] The aggregate time trends indicate that there has been some narrowing in the overall black-white and Hispanic-white test score gaps since the 1970s, but that the gap has remained relatively constant or even grown in the 1990s. There is still a substantial disparity, with black children scoring about 15–25 percent lower than whites on average and Hispanic children about 10 percent lower.[3]

The belief that eliminating racial differences in test score performance would reduce inequality in labor market outcomes is a major motivation for the extensive, multidisciplinary literature aimed at understanding the determinants of children's test scores and the underlying causes of racial/ethnic disparities.[4] A large body of research examines the role of parental characteristics, the early home environment, and school quality in producing cognitive skills. However, this literature has not yet reached a consensus on which inputs increase children's achievement and to what extent, or on the relative contribution of home inputs, school inputs, and family background in accounting for racial/ethnic differences in achievement. A leading candidate for explaining why studies differ so much in their conclusions, even when based on

[1] Robertson and Symons (1990) find that age 7 test scores predict occupational choices, and Currie and Thomas (1999) document their correlation with adult educational and labor market outcomes. These studies are based on data from the British National Child Development Survey.

[2] See also related discussion in Phillips, Crouse, and Ralph (1998), Fryer and Levitt (2004, 2005), and Sec. III of this article. There is, however, some debate over whether test scores widen as children progress through school grades (Ludwig 2003). The timing and magnitude of the widening gap also appear to depend on the test score metric used (Reardon 2007) and on the test measure used (Murnane et al. 2006).

[3] See Jencks and Phillips (1998) and Cook and Evans (2000) for a discussion of trends in scores on National Assessment of Educational Progress (NAEP) tests. Hedges and Nowell (1998, 1999) analyze data from six surveys that include Equal Educational Opportunity Data, National Longitudinal Study of the High School Class of 1972, High School and Beyond, the 1979 National Longitudinal Survey of Youth (NLSY79), the National Education Longitudinal Study of 1988, and NAEP. More recent evidence from the NAEP reported by the National Center of Education Statistics (2005) indicates that the black-white gap stayed constant or grew during the 1990s.

[4] A review of the literature can be found in Todd and Wolpin (2003).

the same data sets, is the wide variety of empirical specifications adopted in the empirical literature (Krueger 1998, 2003; Todd and Wolpin 2003).

Ideally, in analyzing cognitive achievement of children, it would be useful to have data on all past and present home and school inputs as well as information on children's heritable endowments. No data set is that comprehensive, so researchers have had to confront problems of missing or imprecisely measured variables. One approach taken in the literature develops estimators that explicitly recognize the presence of missing inputs. For example, Murnane, Maynard, and Ohls (1981) use school fixed effects to address the problem of missing data on school inputs, under the assumption that children within the same school receive the same inputs. Along similar lines, Rosenzweig and Wolpin (1994) and Altonji and Dunn (1996) adopt a sibling fixed effect approach to address the same problem. Another remedy that is commonly taken when the data lack information on historical input measures is to use a value-added specification that assumes that a previous test score is a sufficient statistic for the missing historical inputs. Finally, some studies implicitly pursue a different strategy of substituting input demand functions in place of missing inputs.[5]

This paper has two main goals: to quantify the impact of home inputs, school inputs, and measured mother's accumulated abilities on children's achievement and to analyze the relative contribution of each factor in accounting for racial/ethnic test score gaps. The measure of mother's abilities that we use, the Armed Forces Qualifying Test (AFQT) score, has been interpreted as a measure of premarket skills because the mothers were aged 15–22 when they took the test (see, e.g., Neal and Johnson 1996). Our main innovation relative to the earlier literature is to implement a cumulative production function for children's cognitive achievement that allows achievement at a given age to depend on the lifetime history of family and school inputs as well as on mother's abilities and unmeasured heritable endowments. Our modeling approach builds on work by Boardman and Murnane (1979), who were the first to formalize a cumulative model of the cognitive achievement production function and to discuss its potential implementation in cross-section and panel data settings. It also extends Todd and Wolpin (2003), which surveyed the literature on estimating production functions for cognitive achievement and discussed the identification assumptions of alternative estimators. This article considers additional estimation approaches and implements alternative estimators with a focus on understanding the sources of racial and ethnic test score disparities.

Our work is complementary to recent work by Cunha and Heckman (2003) that extends the production function framework to incorporate the development of noncognitive skills and their influence on cognitive skill development. They adopt a value-added specification for the joint

[5] See Sec. IV below for a discussion of this approach.

formation of cognitive and noncognitive skills, allowing for measurement error in skills and in home inputs. One interesting finding from their work is that noncognitive skills promote the formation of cognitive skills but not vice versa. They also find evidence for critical skill investment periods during which time the investment needs to be made to be effective, and they demonstrate the importance of early investments. As discussed later in this article, some of the specifications for the cognitive achievement production function that we implement are consistent with the approach of Cunha and Heckman, although we do not explicitly model how noncognitive skills are formed along with cognitive skills.

An important issue in the estimation of the cognitive achievement production function studies is how to select among competing model specifications. We address the model selection problem by applying cross-validation criteria in addition to conventional specification tests. Cross-validation methods find the model that performs best according to an out-of-sample root mean-squared error (RMSE) criterion. They provide a useful alternative to conventional specification testing in situations in which the models being compared are nonnested or when it is not clear which is the preferred null hypothesis model. Cross-validation also seems particularly well suited to our intended use of the estimated model, which is to decompose test score gaps into components attributable to the home environment, school environment, and measured mother's abilities. Specifically, when we use the estimated model to evaluate how much of the minority-white test score gaps would be closed if black or Hispanic children had the same home and school inputs as white children, we essentially perform an out-of-sample forecast. The cross-validation criterion evaluates the reliability of the model in out-of-sample forecasting.[6]

Our analysis samples are drawn from the National Longitudinal Surveys of Labor Market Experience—Children Sample (NLSY79-CS) merged together with school quality data obtained from two main sources: the Common Core Data (CCD) and the American Federation of Teachers (AFT). The strength of the NLSY79-CS data set relative to other available data sets is that it has unusually detailed longitudinal information on children's home environments over the child's entire lifetime and on child achievement as measured by scores on tests that are administered biannually, up to four times per child. The CCD and AFT data are used to derive a time series of school quality, as measured by pupil-teacher ratios and teacher salary.

There is substantial controversy in the educational production function literature over whether pupil-teacher ratios and teacher salaries matter for student outcomes. For example, a widely cited meta-analysis

[6] In a later section, we also discuss potential problems associated with using a theoretical approach such as cross-validation as a tool for model selection.

by Hanushek (1986) concludes that pupil-teacher ratios and teacher salary are not strongly or systematically related to student outcomes. However, Hedges, Laine, and Greenwald (1994) reanalyze the same data, using different methods to synthesize results of the different studies, and conclude that empirical quality measures do matter for student performance. Card and Krueger (1996*b*) and Krueger (2003) argue that the meta-analysis results are sensitive to how estimates from various studies are weighted and whether studies that report more estimates receive more weight. Finally, previously reported evidence on lack of significance is not necessarily relevant to the analysis reported here, as we estimate cumulative specifications of the production function that have never before been estimated.

Using the NLSY-CS data combined with the school quality data, we implement alternative specifications of the production function. The estimates strongly support the notion that skill acquisition is a cumulative process; both contemporaneous and past home inputs are statistically significant determinants of test score outcomes. The school input effects, however, are generally imprecisely measured. Cross-validation criteria as well as conventional model specification tests indicate support for an augmented value-added formulation of the production function. Using our estimates of the cognitive achievement production function parameters, we examine the extent to which home input differences and mother's AFQT score differences can account for racial disparities in test scores among African American, white, and Hispanic children. The empirical results show that equalizing home input levels at the average level observed for white children would close about 10–20 percent of the black-white and Hispanic-white test score gaps (in both math and reading).[7] We also find that the estimated cognitive achievement production function fits well the pattern of rising black-white test score gaps with age as well as differences in test score gap patterns between girls and boys.[8]

This article proceeds as follows. Section II proposes a conceptual framework for modeling the cognitive achievement production function and considers its empirical implementation. Section III describes our data sources and the variables used to represent home and school inputs into the production process. Section IV presents estimates obtained

[7] Our finding that home input gaps are important in accounting for racial test score gaps contrasts with findings reported in recent work by Fryer and Levitt (2004). That paper argues that home input gaps cannot account for black-white test score gaps, because home input gaps remain roughly constant over time whereas test score gaps widen with age. However, their specification assumes that test scores depend only on current home inputs. A cumulative specification that allows for lagged inputs to have an effect can explain a widening black-white test score gap, because a constant home input gap over time implies a widening cumulative gap.

[8] Our work differs from earlier studies, in part, because our specifications allow for unobserved endowment effects and for the cumulative effects of lagged inputs. See Fuchs and Reklis (1994) and Cook and Evans (2000) for different kinds of decompositions using the NAEP data.

under alternative specifications and also presents the cross-validation results. Section V uses the estimated cognitive achievement production function to evaluate the sources of racial disparities in test scores, and Section VI presents conclusions.

II. Alternative Approaches to Modeling and Estimating the Production Function for Achievement

In this section, we lay out a framework for modeling the cognitive achievement production function. It assumes that knowledge acquisition is a cumulative process by which current and past inputs are combined with a child's genetic endowment of mental capacity (determined at conception) to produce a cognitive outcome.[9] Let A_{ija} denote the achievement for child i residing in household j at age a and $\mathbf{Z}_{ija}(a)$ the vector of all inputs applied at any time up until age a.[10] The child's endowed mental capacity is represented by μ_{ij0}. The achievement production function that relates test scores at age a to all prior investments in the child is given by

$$A_{ija} = A_a(\mathbf{Z}_{ij}(a),\ \mu_{ij0}). \tag{1}$$

The empirical implementation of (1) is challenging because heritable endowments are not observed, data sets on inputs are incomplete (i.e., have incomplete input histories or missing inputs), inputs may be chosen endogenously with respect to unobserved endowments, and scores on standardized tests measure achievement with error.

Let T_{ija} be the test score measure that is observed and ε_{ija} a measurement error. Also, let X_{ija} and v_{ija} represent observed and unobserved inputs at age a. To arrive at an empirically implementable specification, assume that the production function is approximately linear in the inputs and in the unobserved endowment and that input effects do not depend on the child's age, but may depend on the age at which they were applied relative to the current age:

$$\begin{aligned} T_{ija} = {} & X_{ija}\alpha_1 + X_{ija-1}\alpha_2 + \cdots + X_{ij1}\alpha_a \\ & + \beta_a\mu_{ij0} + v_{ija}\rho_1 + v_{ija-1}\rho_2 + \cdots + v_{ij1}\rho_a + \varepsilon_{ija}. \end{aligned} \tag{2}$$

In the following discussion, we take this specification to be the "true" structure in the sense that it encompasses most of the common specifications found in the literature. Data limitations have required re-

[9] The production function framework was first formally modeled by Ben-Porath (1967) in the context of an individual decision maker choosing the level of (time and money) resources to devote to human capital investments. It has since served as the basis for much of the literature on skill acquisition in economics. Leibowitz (1974) was the first to extend this conception to home investments in children.

[10] We include in $\mathbf{Z}$ exogenous environmental factors, but for ease of notation do not distinguish them.

searchers to place restrictions on (2). The following specifications and their associated restrictions have been adopted or proposed in previous studies and are implemented in the empirical work of this study.

Contemporaneous specification.—This specification relates an achievement test score to data only on contemporaneous inputs:[11]

$$T_{ija} = X_{ija}\alpha_1 + e_{ija}, \tag{3}$$

where e_{ija} is a residual term that includes the effect of any omitted inputs, lagged inputs (observed and unobserved), endowments, and measurement error. The assumption required to consistently estimate α_1 is that omitted factors, if there are any, be orthogonal to the included input measures.

Cumulative specification with orthogonal endowments and omitted inputs.—This specification expands the contemporaneous specification to include observable lagged inputs, but it maintains the assumption that any omitted inputs and endowments are orthogonal to the included inputs:

$$T_{ija} = X_{ija}\alpha_1 + X_{ija-1}\alpha_2 + \cdots + X_{ij1}\alpha_a + e_{ija}.$$

Fixed-effect specifications.—Fixed-effect specifications provide ways of implementing either the contemporaneous model or the cumulative model in a way that allows input choices to be endogenous with respect to unobserved endowments. Two "fixed-effect" estimators that are prominent in the literature use variation that occurs within families, either across siblings or within children across different ages.

Within-child fixed-effect estimators are feasible when there are multiple observations on achievement outcomes and on inputs for a given child at different ages. Consider differencing the achievement test scores at two ages, a and $a-1$:

$$\begin{aligned} T_{ija} - T_{ija-1} = {} & (X_{ija} - X_{ija-1})\alpha_1 + (X_{ija-1} - X_{ija-2})\alpha_2 \\ & + \cdots + (X_{ij2} - X_{ij1})\alpha_{a-1} + X_{ij1}\alpha_a \\ & + [\beta_a - \beta_{a'}]\mu_{ij0} + e_{ija} - e_{ija-1}. \end{aligned} \tag{4}$$

The parameters of (4) can be consistently estimated under the following assumptions. The first is that the impact of the endowment on achievement must be independent of age ($\beta_a = \beta_{a'}$), so that differencing eliminates the endowment from (4). In that case, orthogonality between input choices and endowments need not be assumed. However, because any prior achievement outcome are known when later input decisions are made, it is necessary to assume that later input choices are invariant

[11] Fryer and Levitt (2004) estimate a version of the contemporaneous model that assumes that inputs do not cumulate and does not allow for endogeneity of inputs.

to prior own achievement outcomes (i.e., that e_{ija-1} is orthogonal to X_{ija}). It is also required that differenced included inputs are orthogonal to differenced omitted inputs or that omitted inputs are age invariant (and are therefore eliminated by the differencing).

The within-family fixed-effect estimator assumes that children of the same parents have a common heritable component. Without loss of generality, decompose the endowment into a family- and a child-specific component, denoted by μ_0^f and μ_0^c. Rewriting (2) yields

$$T_{ija} = X_{ija}\alpha_1 + X_{ija-1}\alpha_2 + \cdots + X_{ij1}\alpha_a + \beta_a\mu_{ij0}^f + \beta_a\mu_{ij0}^c + e_{ija},$$

where the effect of unobservable current and lagged inputs is subsumed into the residual e_{ija}. Consider the estimator in the case of two siblings (i and i') observed at the same age. Differencing the above equation yields

$$\begin{aligned} T_{ija} - T_{i'ja} = {} & (X_{ija} - X_{i'ja})\alpha_1 + \cdots + (X_{ij1} - X_{i'j1})\alpha_a \\ & + [\beta_a(\mu_{ij0}^c - \mu_{i'j0}^c) + e_{ija} - e_{i'ja}]. \end{aligned} \tag{5}$$

Consistent estimation of input effects by ordinary least squares (OLS) requires that inputs associated with any child not respond to the child-specific component of either the own or the sibling endowment. However, input choices may respond to the family component of the endowment. That is, parents who perceive their children to be, on average, of high ability may choose different inputs than other parents, but they are assumed to not make input choices in a way that differentiates with respect to perceived child-specific ability.

Given that achievement is measured for each sibling at the same age, the older child's achievement observation (say child i) will have occurred at a calendar time prior to the younger sibling's observation. Thus the older sibling's achievement outcome was known at the time input decisions for the younger child were made, at the ages of the younger child between the older and younger child's achievement observations. Therefore, consistent estimation of (5) by OLS also requires that input choices not respond to prior sibling outcomes.[12] With regard to omitted inputs, it is necessary to assume that they are invariant across children of the same age or are orthogonal to included inputs.

Value-added specification.—The educational production function literature commonly adopts value-added specifications when data on lagged inputs are missing or incomplete (see, e.g., Tamura 2001). In its most basic form, the value-added specification relates an achievement out-

[12] In essence, this estimation procedure can be justified when intrahousehold allocation decisions are made ignoring child-specific endowments and prior outcomes of all the children in the household (Rosenzweig 1986).

come measure to contemporaneous measures of school and family inputs and to a lagged (baseline) achievement measure:

$$T_{ija} = X_{ija}\alpha + \gamma T_{ij,a-1} + e_{ija}. \tag{6}$$

One can see the restrictions imposed by this form of the value-added specification by deriving it from the unrestricted model given in (2). Subtract $\gamma T_{ij,a-1}$ from both sides of (2) and collect terms to get

$$\begin{aligned} T_{ija} = {} & X_{ija}\alpha_1 + \gamma T_{ij,a-1} + X_{ija-1}(\alpha_2 - \gamma\alpha_1) + \cdots + X_{ij1}(\alpha_a - \gamma\alpha_{a-1}) \\ & + (\beta_a - \gamma\beta_{a-1})\mu_{ij0} + (e_{ija} - \gamma e_{ij,a-1}), \end{aligned} \tag{7}$$

where

$$\begin{aligned} e_{ija} - \gamma e_{ij,a-1} = {} & v_{ija}\rho_1 + v_{ija-1}(\rho_2 - \gamma\rho_1) + \cdots + v_{ij1}(\rho_a - \gamma\rho_{a-1}) \\ & + \varepsilon_{ija} - \gamma\varepsilon_{ija-1}. \end{aligned}$$

For (7) to reduce to (6), we require the following assumptions:[13]

i. For all k, $\alpha_k = \gamma\alpha_{k-1}$, which implies that coefficients associated with observed inputs geometrically decline (presumably) with number of years since application of the input. Also, the rate of decline is assumed to be the same for each input.
ii. If there are omitted inputs, then condition i also has to hold for omitted inputs ($\rho_k = \gamma\rho_{k-1}$) and the contemporaneous omitted input v_{ija} has to be uncorrelated with included inputs and with the baseline test score. Alternatively, it could be assumed instead that all omitted inputs (current and lagged) are uncorrelated with included inputs and with the baseline test score.
iii. The impact of the endowment (μ_{ij0}) geometrically declines at the same rate as input effects, that is, $\beta_a = \gamma\beta_{a-1}$.

For the OLS estimator of α_1 to be consistent, ε_{ija} must also be serially correlated, with the degree of serial correlation matching the rate of decay of input effects (so that $\varepsilon_{ija} - \gamma\varepsilon_{ija-1}$ is an independently and identically distributed [iid] shock). If this condition does not hold, then baseline achievement, T_{ija-1}, will necessarily be correlated with its own measurement error. The iid assumption could be relaxed to allow, for example, an AR(1) structure on the residuals, provided that additional lagged test score measures are available that could be used as instruments.

In summary, strict value-added models that include lagged test scores and current inputs impose strong assumptions on the pattern of the coefficients associated with the inputs. As noted above, the effect of inputs applied at one point in time has to decline with the age of the

[13] See also Boardman and Murnane (1979) for a related discussion of these conditions. Equation (7) is the well-known Koyck transformation.

child. For example, the effect of reading to a child at ages 3–5 would be assumed to have a greater impact on reading scores at age 8 than at age 10. In the empirical work reported in Section IV, we find some support for the pattern of declining coefficients with age; however, statistical tests reject the strict value-added formulation. The value-added model also imposes a strong assumption on the orthogonality of contemporaneous omitted inputs and lagged test scores, which may not hold if omitted inputs are correlated over time.

Value-added plus lagged inputs specification.—When data are available on historical input measures, then assumption i can be relaxed. A value-added specification that does not impose assumption i would include as additional regressors data on lagged inputs. As noted in the introduction, recent work by Cunha and Heckman (2003) and Cunha et al. (2006) adopts a value-added specification to jointly model the formation of cognitive and noncognitive skills. In terms of our notation, their specification can be represented as

$$\begin{pmatrix} T_{ijt+1}^{N} \\ T_{ijt+1}^{C} \end{pmatrix} = A_t \begin{pmatrix} T_{ijt}^{N} \\ T_{ijt}^{C} \end{pmatrix} + B_t X_{ija} + \begin{pmatrix} e_{ijt}^{N} \\ e_{ijt}^{C} \end{pmatrix},$$

where T_{ijt+1}^{N} denotes a measure of noncognitive skills and T_{ijt+1}^{C} a measure of cognitive skills, and X_{ija} are current inputs. If one were to focus only on the cognitive test score measure equation and to substitute repeatedly for the right-hand-side T_{ijt}^{N} variables, then a value-added plus specification would be obtained. In that sense, the model they implement is consistent with the value-added plus specification.

III. Data

As described in Section II, the data requirements for implementing the cumulative specifications of the cognitive achievement production function are demanding. A researcher needs a history of inputs, beginning at the child's conception. In addition, to account for unobserved endowments, one needs multiple observations on achievement measures, either for siblings at the same ages or for the same child at different ages. Although there does not exist a data set that satisfies all these requirements, the NLSY79-CS comes close.

The NLSY79-CS is a sample of all children ever born to the women respondents of the NLSY79. The NLSY79 is itself a nationally representative sample of individuals who were aged 14–21 as of January 1, 1979, with significant oversamples of blacks and Hispanics. The survey collects extensive information about schooling, employment, marriage, fertility, income, assets, alcohol and drug use, participation in public programs, and other related topics, many as event histories. For example, employment events are known up to the week, marriage and fertility events to the day, and school enrollment to the month. This enables the researcher to create an almost complete life history for each respondent for many important events dating back to age 14.

Beginning with the 1986 interview, separate questionnaires were developed to collect information about the cognitive, social, and behavioral development of the children of the NLSY79 respondents. Questionnaires were administered to the women (mothers) of the children as well as to the children themselves. These interviews have been conducted biannually since 1986. By 2000, over 11,000 children were interviewed. Approximately 28 percent of the children in 2000 were African American, 19 percent Hispanic, and the rest mostly white.

Cognitive achievement measures.—Our analysis restricts attention to two cognitive tests that were administered to all children starting at age 5: the Peabody Individual Achievement Test in mathematics (PIAT-M) and the Peabody Individual Achievement Test in reading recognition (PIAT-R). The PIAT tests are designed to measure cognitive achievement. They were administered each year of the survey, and many of the children in the sample have two or more scores. Completion rates for the PIATs have been around 90 percent.

In our estimation, we use both raw test scores and age-normed percentile scores. The PIAT math and reading tests were designed so that the raw score provides an absolute measure of achievement capturing gains in knowledge over time as additional input investments are made in a child. In particular, test questions are ranked in terms of difficulty, and children are asked questions as long as they are able to answer them. In principle, a first grader might even be asked high school–level math questions if he or she keeps answering questions correctly. The raw score therefore captures an absolute measure of knowledge (numbers of questions answered correctly) that is comparable across children of different ages.[14] The age-normed percentile score provides a relative measure of performance that depends only on ordinal test score rankings and not on actual numbers answered correctly. Some researchers find so-called metric-free test score measures to be more meaningful (Reardon 2007). Clearly, the percentile score is derivable from the raw score, and the use of one versus another for estimating an education production function essentially amounts to a functional form assumption. If the estimation of the relationship between test score outcomes and school and home inputs were fully nonparametric, then either raw scores or percentile scores (a known transformation of raw scores) could be used and would recover, up to the transformation, the same pro-

[14] The PIAT testing procedure starts at a point deemed appropriate for the child's grade level and then moves backward in difficulty until five questions are answered correctly, which establishes a basal score. Starting from the basal, the child then answers questions until five out of seven are answered incorrectly, which establishes the ceiling. The raw score equals the basal score plus the number of questions that were answered correctly between the ceiling and the basal scores. This scoring method assumes that children would have correctly answered the questions below the basal score.

TABLE 1
Descriptive Statistics: Means and Standard Deviations (in Parentheses)

	White	Black	Hispanic
Piat math percentile (ages 6–13)	61.9	42.4	47.2
	(24.5)	(25.2)	(26.4)
Piat reading percentile (ages 6–13)	60.1	44.2	48.9
	(26.5)	(27.2)	(28.9)
Piat math raw score (ages 6–13)	42.0	34.9	36.0
	(14.2)	(13.7)	(14.3)
Piat reading raw score (ages 6–13)	40.5	34.5	36.0
	(14.0)	(12.4)	(14.0)
Current home score (ages 6–13)	104.8	88.9	91.3
	(20.6)	(23.8)	(24.5)
Average home score at ages 3–5	123.2	106.0	108.3
	(16.9)	(23.3)	(25.6)
Average pupil-teacher ratio (over all of the child's school years)	18.3	18.1	19.8
	(2.4)	(2.4)	(3.3)
Average teacher salary (1989 $) (over all of the child's school years)	31,661	29,624	32,218
Child age (in months)	113.8	113.4	111.9
	(24.6)	(24.9)	(24.4)
Birth weight (ounces)	120.8	112.0	118.9
	(20.2)	(21.7)	(20.7)
Percentage first born	48.3	33.9	40.3
Percentage second born	35.8	37.8	34.5
Percentage with mother's age at birth:			
Less than 18 years	1.1	1.5	1.7
18–19 years	5.0	10.9	11.0
20–29 years	82.0	80.7	80.1
Mother's years of schooling	13.1	12.4	11.7
	(2.3)	(2.0)	(11.7)
Mother's AFQT percentile score	52.4	20.4	25.6
	(26.0)	(17.7)	(22.1)
Observations	3,802	2,403	1,495

duction function. Owing to data limitations, a fully nonparametric estimation is not feasible. We therefore use both raw scores and percentile scores to examine sensitivity to the metric used. As discussed in Section IV, the results are not overly sensitive to the metric for the two metrics we examined. A separate consideration is that we compare children by age rather than by grade, because blacks have higher rates of grade retention, which would tend to understate achievement gaps if comparisons were made by grade.[15]

Table 1 shows the average PIAT math and reading scores by race/ethnicity. The average gap in black-white percentile scores is 19.5 and the average gap in Hispanic-white percentile scores is 14.7. The standard deviation is comparable across groups. Figures 1*a* and *b* plot the average PIAT reading and math percentile test scores by age, by gender, and by

[15] See a discussion on this point in Hanushek and Rivkin (2006).

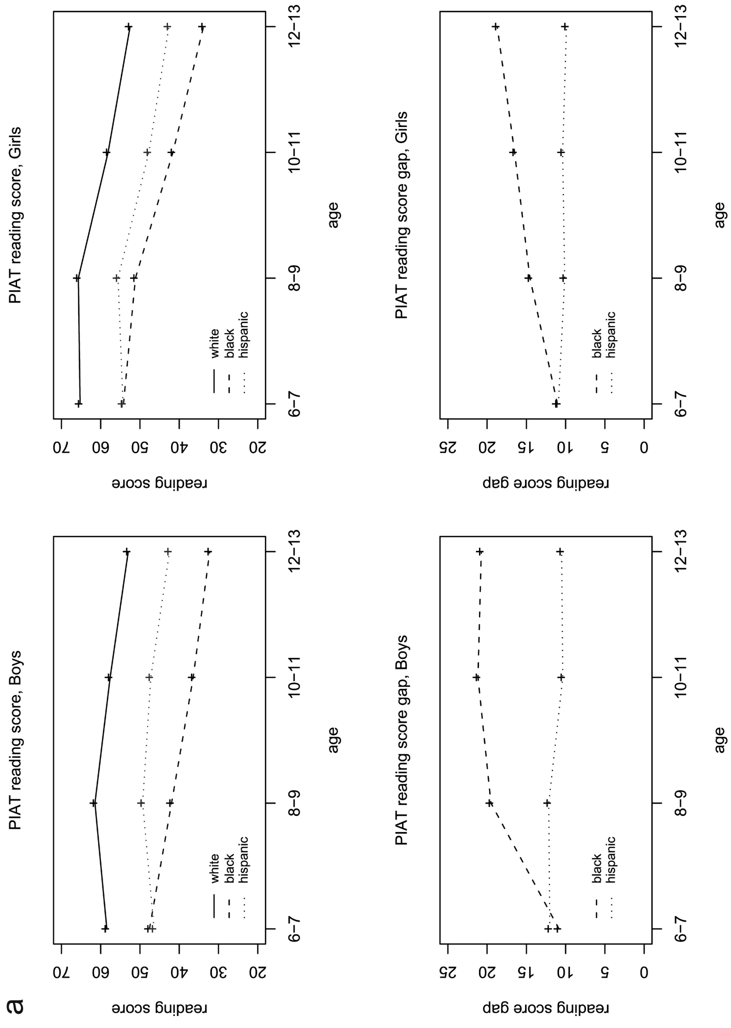
race/ethnicity.[16] The lower panel shows the black-white and Hispanic-white gaps. At age 6, the gap in reading scores for both black and Hispanic children relative to white children is 10 percentile points. The gap remains roughly constant with age for Hispanic children but widens for black children, particularly for boys over ages 6–9. By ages 12–13, the gap for both black boys and girls is approximately 20 percentile points. For math test scores, there is also a sizable gap even at age 6 between whites and minorities. As with reading, the black-white test score gap for boys widens markedly over ages 6–9. Unlike reading, however, the gap for black boys remains approximately constant from age 9 to age 12. Reardon (2007) notes a similar pattern for the widening of black-white reading and math scores for boys. For girls, there is a more pronounced widening of the black-white reading gap than the math gap. The white-Hispanic gap in the reading score exhibits some widening, then convergence for boys, and some convergence and then widening for girls. By ages 12–13, the math gap for black boys is much greater than for Hispanic boys, whereas the gaps for black and Hispanic girls are similar. Figure A1 presents plots analogous to figure 1 for the raw test scores instead of the percentile scores.

Home input measures.—The NLSY79-CS includes a battery of questions about the home environment of the child called the Home Observation Measurement of the Environment—Short Form (HOME-SF).[17] The HOME-SF consists of four different instruments that depend on the age of the child: ages 0–2, 3–5, 6–9, and 10 and above. The instrument is self-administered to the mother of the child. A second version is filled out by the interviewer. Researchers can use either the answers to individual items or scales provided in the public-use files. The total raw score is a simple summation of responses (modified so each has a {0, 1} domain) of individual items.

Some of the items in the home scale can be directly linked to cognitive achievement in the sense that they are related to learning-specific skills. For example, mothers of children under the age of 10 are asked how often they read stories to their child, and mothers of children between the ages of 3 and 5 are asked whether they help their child to learn numbers, the alphabet, colors, or shapes and sizes. Other items are not as clearly tied to cognitive achievement but may be thought of as contributing to an environment conducive to learning. For example, mothers are asked how many books the child has, whether the family en-

[16] The survey is a biennial survey, so children are typically interviewed at even ages 6, 8, 10, and 12 or at odd ages 7, 9, 11, and 13. For this reason, we report in the figure averages over age categories 6–7, 8–9, 10–11, and 12–13.

[17] As the name suggests, the short form is a modification of a version that is about twice as long. The HOME was created by Caldwell and Bradley (1984). Some parts of the shortened version used in the NLSY79-CS were created by them, and all were reviewed by them. The HOME-SF is widely used, and there exists considerable research on the validity and reliability (see the citations in the 1996 *Users Guide*).


a
PIAT reading score, Boys
PIAT reading score, Girls
PIAT reading score gap, Boys
PIAT reading score gap, Girls
reading score
reading score gap
age
white
black
hispanic
6–7
8–9
10–11
12–13

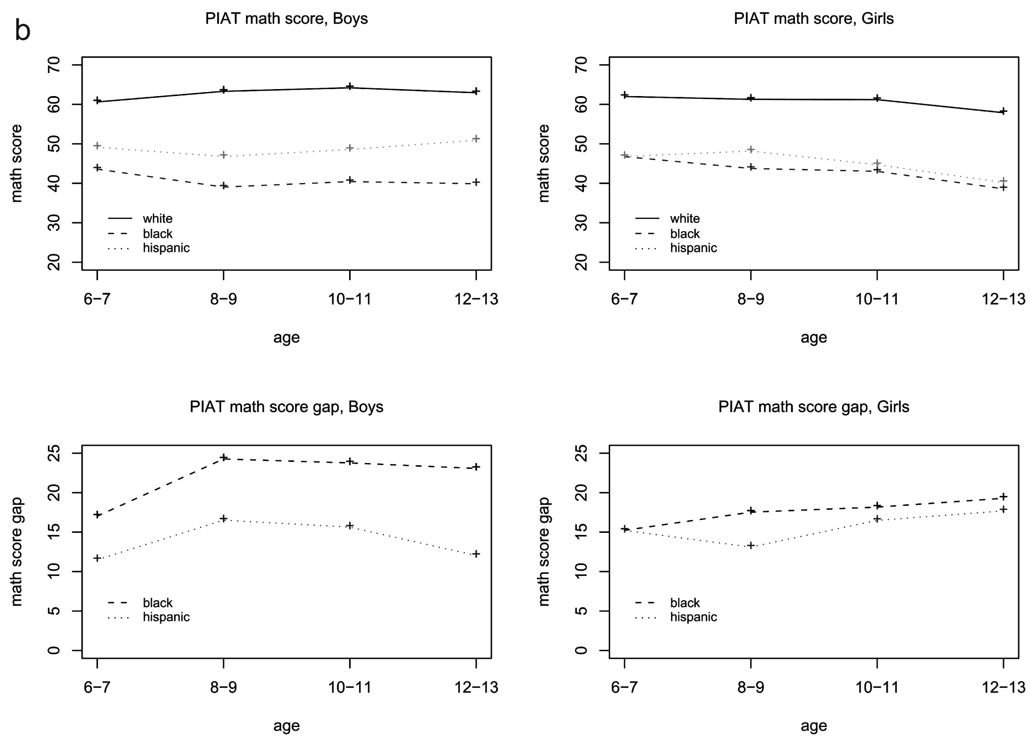


Figure 1.—Comparison of PIAT percentile scores by race/ethnicity: *a*, reading; *b*, math

courages the child to start and keep doing hobbies, whether the child has access to a musical instrument, and whether the family takes the child to museums or theatrical performances. About two-thirds of the items in the home scale are based on mother self-reports of her own and her child's activities and about one-third of the items correspond to interviewer observations about the child's home environment.

In the empirical work reported below, we use the home scale provided in the public-use files as our measure of the time and goods inputs provided in the home. As described in Section II, we consider both current and historical home inputs as potential determinants of current test scores. To get a better idea of what the home scale measures, tables A1–A3 in the Appendix compare the average scores by race/ethnicity for the individual items of the cognitive home scale for children in different age ranges.[18] The average scores for the African American and Hispanic mothers tend to be similar and tend to be lower than the scores for white mothers for most of the individual items. The differences are particularly notable for the questions related to number of books in the child's possession, the number of times the mother reads to the child, and the teaching activities the mother engages in with the child. For example, 94 percent of white mothers report that their age 3–5 toddler has 10 or more books in comparison with 57 percent of black mothers and 63 percent of Hispanic mothers. The difference in book ownership persists for children in all the age ranges. Seventy percent of these same mothers report reading stories to their toddler at least three times a week, in comparison with 40 percent of black mothers and 44 percent of Hispanic mothers. Sixty-six percent of black mothers and 70 percent of Hispanic mothers report teaching their age 3–5 child numbers in comparison with 78 percent of white mothers. For older children aged 6–9, 61 percent of white children receive special lessons or participate in organizations that encourage sports, arts, dance, or drama, compared to 41 percent for black children and 39 percent for Hispanic children. The items of the home scale based on interviewer observations also show some differences by race/ethnicity, but they tend to be smaller than the differences observed on the self-report items. Thus examination of the individual items of the home scores reveals some stark racial/ethnic differences for children in all the age ranges, especially for the items that are self-reported by the mother related to books, reading, and teaching activities.

Table 1 shows the average home score for our analysis samples, for ages 6–13 and for ages 3–5.[19] The average white home score at ages 3–5 is 13 percent higher than the average black home score and 12 percent higher than the average Hispanic home score. At ages 6–13, the white

[18] The questions that are asked of the mother differ slightly across different age ranges.

[19] We distinguish these age groups because the questions that constitute the home score scale are substantially different for 3–5-year-olds (see app. tables A1–A3 for a list of the questions).

score is 15 percent higher than the black score and 13 percent higher than the Hispanic score. Figure 2 plots the current home score by age, gender, and race/ethnicity. The plots show that the gap in home scores (relative to whites) is similar for blacks and Hispanics and declines somewhat with age. The plots by gender show that black boys have slightly lower home scores than Hispanic boys, but the reverse is true for girls. A comparison of the home scores for girls and boys shows that for all the race/ethnicity subgroups, girls tend to have higher home scores than boys.

Maternal characteristics.—Information is available on mother's completed schooling, and it is updated in each year in which the mother attended school. Because some women return to school after having children, both within-family and within-child estimators can be used to estimate maternal schooling effects on children's achievement (see Rosenzweig and Wolpin 1995). Women with higher school attainment presumably have more knowledge to transmit to their children or may be better teachers. A comparison of mothers' schooling levels by race/ethnicity shows that white mothers have the highest average years of schooling (13.1), African American mothers the second-highest (12.4), and Hispanics the lowest (11.7).

In addition to schooling, the NLSY79 also contains a measure of mothers' accumulated abilities, their score on the AFQT.[20] A measure of mother's abilities is a potentially important factor in the production of cognitive skills in children. As seen in table 1, the AFQT score for white mothers is close to the median (52.4), whereas the average percentile rank for African American and Hispanic mothers, 20.4 and 25.6, is much lower. The white-black and white-Hispanic differences are magnified when the average is taken over child observations rather than over mothers, because lower-AFQT mothers have relatively more children.

The NLSY79 includes only limited information about fathers, and identifying the biological father is problematic. Although the public-use data include a variable indicating presence of the biological father in the household, the variable in many cases is missing. In preliminary analyses, we did not find the presence of the father in the household to be an important determinant of test scores. However, to the extent that the effect of the father present in the household is constant over the observation period, it would be captured by the unobserved heterogeneity term when we estimate fixed-effect mother, fixed-effect child, and value-added models.

Child characteristics.—In addition to standard information on race and gender of the child, the NLSY79-CS also contains information on other

[20] We use the 1989 created variable percentile AFQT score. The AFQT is a summary score based on tests of word knowledge, paragraph comprehension, verbal and math knowledge, and arithmetic reasoning.

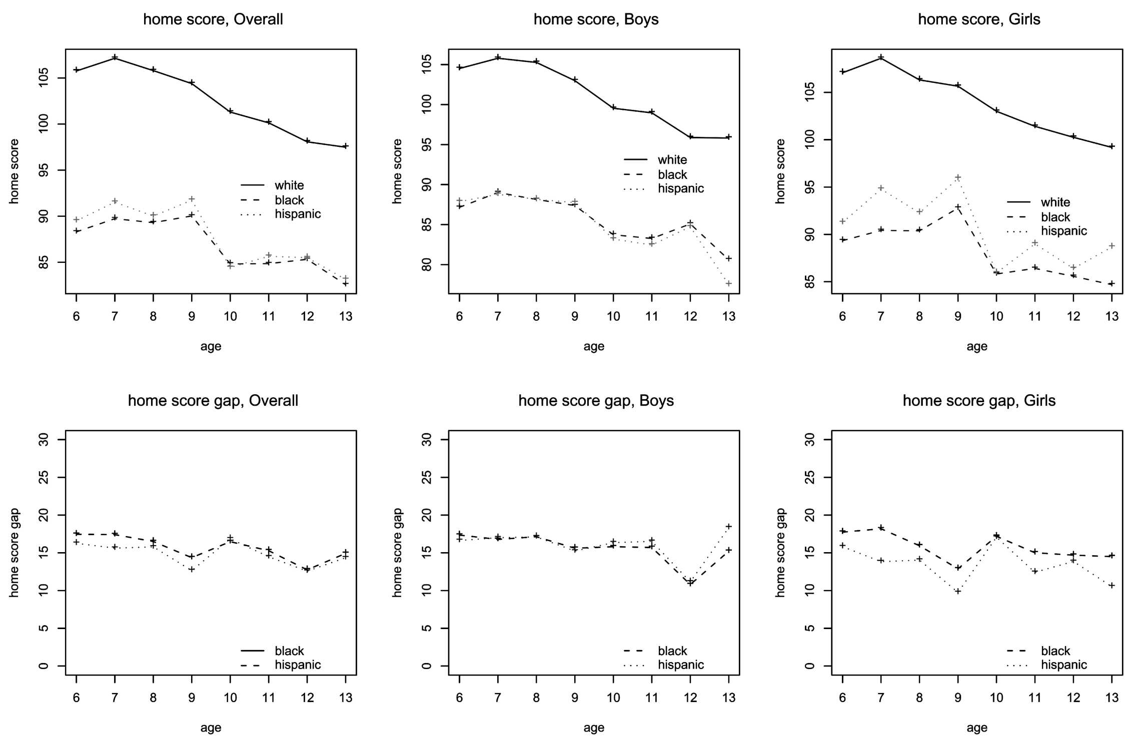


Figure 2.—Comparison of current home score by age by race/ethnicity

characteristics that are potential determinants of a child's cognitive achievement, such as birth order and birth weight (see, e.g., Rosenzweig 1986). As shown in table 1, African American children have, on average, lower birth weight than white or Hispanic children. The disparity of about 6–8 ounces is due either to biological factors or to differences in prenatal investments. In addition to having, on average, higher birth weight, white children are more likely to be first- or second-born, because white women have fewer children than black or Hispanic women.

School inputs.—As noted in the introduction, the strength of the NLSY79-CS data relative to other available data sets is the detailed information on the home environment that spans a child's entire lifetime. The major weakness of the NLSY79-CS is lack of information on schools, which requires supplementing the data set with school data from other sources. Implementing the cumulative model described by equation (1) requires both contemporaneous and historical data on school inputs.[21] It would be preferable to measure school inputs at the same level as home inputs, that is, at the level of the child's classroom, but to our knowledge there does not exist a data set with such detailed information on both the home and school environments. Choosing a data set for analysis requires trade-offs between the level of detail of observation on the home and school environments. Rather than omit school quality from the analysis, we have opted to include school quality even though, in our data, they can be measured only at the state and county levels.[22]

We obtained school data from three sources: the Common Core Data, the School and Staffing Survey, and the American Federation of Teachers. The CCD provides information on all public schools and on the characteristics of students at both the school and district levels. In the CCD, schools report the number of full-time-equivalent teachers and the number of pupils enrolled, which we use to calculate pupil-teacher ratios for each school. Because elementary grades and upper-level grades are usually offered in separate schools, we calculate separate pupil-teacher ratio averages for grades 1–6 and grades 7–12. We constructed both county-level and state-level pupil-teacher ratio variables, which we merged with the NLSY79-CS data.[23] We also obtained a series of average teacher salaries by state, for the years 1984–2001, from the AFT.

The schooling inputs on which we focus in the analysis are pupil-teacher ratios and teacher salaries. We also estimated specifications using data on teacher's education, teacher experience, hours/weeks spent

[21] County and state of residence are available at each survey round of the NLSY79 respondents (and their children) and can be obtained as a restricted data file from the Bureau of Labor Statistics.

[22] Nonetheless, school quality measured at similar levels of aggregation has been shown to be correlated with labor market outcomes. See, e.g., Card and Krueger (1992), Betts (1995), and Tamura (2001).

[23] For confidentiality reasons, the NLSY data contain information only on the child's county of residence and not on the school attended.

teaching math and English (separately), and teacher certification.[24] These variables do not appear in the final specifications as inputs because estimates of their effects were never precise. We therefore adopted a more parsimonious specification that includes only two conventional measures of school inputs: pupil-teacher ratios and teacher salaries.

Table 1 shows average pupil-teacher ratios and average teachers' salaries for white, African American, and Hispanic children, where the average is taken over the child's school history for the years in which the school input measures are available.[25] Although, historically, African American children attended schools that were of much lower quality than white children, there has been substantial convergence in empirical measures of schooling quality over time. Boozer, Krueger, and Wolken (1992) note that in 1970 the pupil-teacher ratios in schools attended by black children were on average 11 percent higher than in schools attended by whites, but by 1990 there was no difference.[26] In our schooling data, the average pupil-teacher ratios are lowest for African American children (18.1) and highest for Hispanic children (19.8). Teacher salaries, on average, are highest for Hispanic children ($32,218) and lowest for African American children ($29,624).

IV. Empirical Results

As described in Section II, a benefit of rich longitudinal data is that they enable estimation of more general specifications that accommodate the presence of unobserved endowments and input choices that are endogenous with respect to those endowments. Here, we estimate all the specifications described in Section II: the contemporaneous specification, the cumulative specification with orthogonal endowments and unobserved inputs, within-child and sibling fixed-effect specifications, a value-added specification, and the value-added plus specification (described in Sec. II, which includes lagged inputs). As previously noted, these specifications impose different sets of restrictions on the most general model, given in (2).

[24] These data are available from the School and Staffing Survey for a random sample of one in seven schools, which we aggregated to the state and county levels.

[25] We attempted to construct separate contemporaneous and lagged average measures of school inputs (as we do with the home inputs), but, perhaps because of the higher level of aggregation, there was substantial colinearity and we were unable to obtain precise estimates of their separate effects. Therefore, we use cumulative measures that depend on the child's schooling history.

[26] See also Card and Krueger (1992) for evidence on the convergence of schooling quality in black and white schools over the last century and an analysis of the effects of convergence on earnings. Donohue, Heckman, and Todd (2002) study the sources of convergence in the South over the 1911–60 time period.

TABLE 2
AVAILABLE LAGGED HOME INPUT MEASURES AT DIFFERENT AGES AND CORRESPONDING COEFFICIENTS

Age at Test Score Measure	Age at Current Home	Age at One-Period-Lagged Home	Age at Two-Period-Lagged Home	Age at Three-Period-Lagged Home	Age at Four-Period-Lagged Home
6–7	6–7 (α_0)	3–5 (α_{11})	. . .	. . .	. . .
8–9	7–8 (α_0)	6–7 (α_{12})	3–5 (α_{21})	. . .	. . .
10–11	10–11 (α_0)	8–9 (α_{12})	6–7 (α_{22})	3–5 (α_{31})	. . .
12–13	12–13 (α_0)	10–11 (α_{12})	8–9 (α_{22})	6–7 (α_{32})	3–5 (α_4)

A. Estimating Equations

The cumulative model presented in Section II relates test scores to current and lagged home and school inputs. Children are interviewed approximately every two years, so the coefficients on the inputs can be interpreted as an effect of applying the input over a two-year interval. Table 2 shows the age at the time of measuring the current and lagged home inputs and which measures are available at different ages. We specify the dependence of test scores on lagged home and school inputs as

$$\begin{aligned}T_a = {}& \alpha_0 \text{Home}_a + \alpha_{11}\text{LagHome}_a^{-1} I(\text{ages } 6\text{–}7)\\ &+ \alpha_{12}\text{LagHome}_a^{-1} I(\text{ages } 8\text{–}13)\\ &+ \alpha_{21}\text{LagHome}_a^{-2} I(\text{ages } 8\text{–}9) + \alpha_{22}\text{LagHome}_a^{-2} I(\text{ages } 10\text{–}13)\\ &+ \alpha_{31}\text{LagHome}_a^{-3} I(\text{ages } 10\text{–}11) + \alpha_{32}\text{LagHome}_a^{-3} I(\text{ages } 12\text{–}13)\\ &+ \alpha_4\text{LagHome}_a^{-4} I(\text{ages } 12\text{–}13)\\ &+ \delta_1 \text{PTRAvg}_a \times S_a + \delta_2 \text{TchSalAvg}_a \times S_a\\ &+ \varphi_1 \text{AFQT} \times I(\text{ages } 6\text{–}7) + \varphi_2 \text{AFQT} \times I(\text{ages } 8\text{–}9)\\ &+ \varphi_3 \text{AFQT} \times I(\text{ages } 10\text{–}11) + \varphi_4 \text{AFQT} \times I(\text{ages } 12\text{–}13)\\ &+ \gamma X_a + \beta_a \mu + \varepsilon_a, \qquad (8)\end{aligned}$$

where T_a represents the test score at age a. The notation $I(\text{ages } xx)$ is an indicator for whether the child's age at the time of the test is in a given age range. Home_a represents the contemporaneous home input and LagHome_a^{-k} represents the kth-period lagged home input. We allow the coefficient on the home input measure to differ when the lag corresponds to the home input measured at ages 3–5, because the battery of questions that constitute the home scale in the early ages is different (see Sec. III). For this reason, α_{21} is not restricted to equal α_{22} and α_{31} is not restricted to equal α_{32}.

The variables PTRAvg_a and TchSalAvg_a are the average of the school quality variables (pupil-teacher ratios and teacher salary) taken over the years in which the child attended school. The average takes into account the effects of state-to-state migration, effects of changing schools (e.g., from elementary to middle to high school), and intertemporal variation. These all provide sources of within-child variation in school quality. The average school quality is multiplied by the number of years attended (S_a) to get the cumulative effect of having been exposed to that level of school quality. Although the school quality variables could have been treated symmetrically with the home input variables (by including separate lags instead of the average), this was impractical because of colinearity in the lag school measures.[27]

We also include in the specification the mother's AFQT score measure under the presumption that mothers with a higher set of skills have a technological advantage in the production of children's achievement. The coefficient associated with AFQT is allowed to vary depending on the child's age for two reasons: (i) to capture the cumulative effect of AFQT and (ii) to allow for heterogeneity in how AFQT affects test scores with respect to the child's age. For example, mother's AFQT may be more relevant for the cognitive achievement of older children. Finally, the variable X_a represents additional covariates in the specification that include birth weight, indicators for whether child is the first- or second-born child, indicators for mother's age at the time of birth, the child's age in months and its square, as well as separate indicator variables for each age and the mother's schooling level.

As noted in Section II, one of the major difficulties in estimating the cognitive achievement production function is how to account for missing data on inputs. There is no completely satisfactory solution to this problem. One possible approach is to assume that missing inputs are orthogonal to included inputs and therefore do not bias the estimation of their associated coefficients. Another approach is to assume that the effects of omitted inputs are constant over time or constant over children within a family, in which case the effect of unobserved inputs would be eliminated by within-child and within-family estimators. Alternatively, under the assumptions discussed in Section II, in relation to value-added models, if the coefficients associated with omitted inputs decline at a geometric rate that matches that of the coefficients on included inputs, then the effect of omitted inputs would be eliminated in a value-added specification.

A third approach to addressing the problem of omitted inputs is to specify input demand equations that express the missing inputs as functions of current and past family income, prices, and preference shocks and to substitute the input demand equation in place of the missing input in the production function. We next illustrate this third approach using a simplified example.

[27] The school quality measures are state-level averages, measured separately for primary, middle, and secondary schools.

Assume that parents (with one child) maximize utility that depends on consumption and child achievement. They can purchase three different inputs, X_1, X_2, and X_3, to produce achievement at prices p_1, p_2, and p_3. We assume that the X's may also be direct sources of utility. The problem solved by parents is

$$\max_{X_1, X_2, X_3} U(A(X_1, X_2, X_3), X_1, X_2, X_3, C, \eta_1, \eta_2, \eta_3)$$

$$\text{subject to } C + p_1 X_1 + p_2 X_2 + p_3 X_3 = y,$$

where A is the child achievement production function, y represents family income, C is consumption, and η_j, $j = 1, \ldots, 3$, are family-specific unobserved preference shifters (e.g., to the marginal utility of inputs, e.g., the utility of reading to a child). The model requires as many unobserved preference shifters as there are inputs to give positive probability to observing all the possible combinations of input choices. The price of consumption has been normalized to one. From this model, one can derive input demand equations that express input choices as a function of prices, family income, and preference shocks:[28]

$$X_j = \gamma_0^j + \gamma_1^j p_1 + \gamma_2^j p_2 + \gamma_3^j p_3 + \gamma_4^j y + \nu_j(\eta_1, \eta_2, \eta_3).$$

Assume that the production function for measured achievement is given by

$$\tilde{A} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon,$$

where ε represents measurement error. If data on X_1 and X_2 are available but data on X_3 are missing, then one approach for dealing with the potential omitted-variables bias is to substitute the linear input demand equation for X_3, which gives a hybrid specification of the production function:

$$\tilde{A} = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_1 + \tilde{\alpha}_2 X_2 + \tilde{\alpha}_3 \gamma_0^j + \tilde{\alpha}_3 \gamma_1^j p_1 + \tilde{\alpha}_3 \gamma_2^j p_2 + \tilde{\alpha}_3 \gamma_3^j p_3$$
$$+ \tilde{\alpha}_3 \gamma_4^j y + (\tilde{\alpha}_3 \nu_j + \varepsilon),$$

where the terms in parentheses constitute the residual.[29] If prices are constant across all the observations, then their effect would be absorbed into the model intercept. This hybrid specification now explicitly accounts for the missing input X_3. However, applying OLS to the estimation of the hybrid specification is problematic because the shocks in the input demand equations would generally be correlated ($E(\nu_j \nu_k) \neq 0$), implying a nonzero correlation between the observed included inputs and the error term. For this reason, there is no a priori reason to

[28] We denote the dependence of the unobservable determinants of input levels on the shocks with the notation $\nu_j(\eta_1, \eta_2, \eta_3)$.

[29] See Rosenzweig and Schultz (1982) for discussion of the interpretation of hybrid production function parameters.

expect that the bias in estimating the technology parameters α_1 and α_2 would be smaller under the hybrid specification than under the original specification that includes X_1 and X_2 and simply omits X_3. One case that would yield consistency of the hybrid model estimators for the production function parameters is when the input demand functions depended only on observables.

In the empirical work, we consider specifications of the form (8) as well as a hybrid specification that adds cumulative family income over the child's lifetime to proxy for omitted inputs.[30] We also include race and gender in the hybrid model to allow family preferences for achievement to potentially vary by race and child gender.

B. *Estimated Production Function Coefficients*

Table 3 reports the estimated production function coefficients for PIAT math and reading percentile test score measures. We estimated the production function separately for math and reading scores, though in many cases the patterns of estimated coefficients are similar across the two scores. In the table, each column reports coefficient estimates for a different model specification, where we implement all the specifications that were described in Section II. Appendix table A4 reports analogous results for the "hybrid" specifications, which include as additional variables cumulative family income, race, and gender. Tables A5 and A6 (included in the online Appendix) report the estimated coefficients when the raw test score measure is used instead of the percentile measure. In addition to the regressors shown in the tables, the model includes the additional regressors specified in the table notes.

As discussed in Section III, the contemporaneous specification places strong restrictions on the production technology but is less demanding than other specifications in terms of data requirements. Under the null that the contemporaneous model is correctly specified, test scores are a function only of current input measures. The first column of table 3, labeled Contemporaneous, presents the estimated input coefficients for the math test score (panel A). The coefficient associated with the current home input is positive and statistically significantly different from zero, as are the coefficients corresponding to the mother's AFQT score. A 16-point increase in the home score (which would close the gap in average home scores between blacks and whites) would lead to an increase of 2.7 in the percentile score. The corresponding estimates for the reading percentile score, shown in panel B, exhibit almost identical patterns. The coefficients associated with the pupil-teacher ratio and the teacher salary usually have the expected signs for the math percentile

[30] The same approach applied to a cumulative specification would substitute the input demand equations for both contemporaneous and lagged home inputs. These would depend on current and lagged family income.

scores but tend to be statistically insignificantly different from zero.[31] A comparison of table 3 and table A4 shows that the estimated home input effects are very similar across the hybrid and nonhybrid specifications. A comparison with tables A5 and A6 in the online Appendix shows that the estimated home input and mother's AFQT effects remain precisely estimated when raw scores instead of percentile scores are used.

A straightforward test of the contemporaneous specification that is feasible when historical data on inputs are available is to include the historical input measures in the specification and check whether their associated coefficients are significantly different from zero. The second column of numbers in each table augments the contemporaneous specification to include lagged data on home inputs.[32] The estimated effect of the current home input falls substantially when lagged inputs are included, which suggests that omitting historical measures leads to an overstatement of the immediate impact of a unit increase in home input. At the same time, it leads to an understatement of the impact of a unit increase in the home score that is sustained over an extended time period, because it neglects that home inputs have lasting effects. For example, the estimates for the reading specification (panel B of table 3, col. 2) imply that a unit increase in the home score at ages 3–5, 6–7, and 8–9 increases the PIAT reading test score at ages 10–11 by $0.11 + 0.05 + 0.05 + 0.06 = 0.27$, which is 58 percent larger than the effect implied by the contemporaneous specification. The estimated coefficients on all but one of the lagged inputs (the third lag at ages 12–13) are positive and statistically significantly different from zero, which is evidence against the contemporaneous specification. As seen in table 3, mother's AFQT is an important determinant of children's test scores. The pattern of the estimated coefficients observed for both math and reading scores is that the importance of mother's AFQT increases as children get older. The increasing importance of mother's AFQT may reflect either its cumulative effect or that children benefit more from having higher-AFQT mothers when they are older.

The third column of table 3 (labeled Value Added (1)) implements the strict value-added specification that includes a one-period-lagged test score along with contemporaneous home and school inputs. The coefficients on the home input variable and on the lagged test score measures are statistically significantly different from zero and have the expected sign (positive). Panels A and B of table A4 calculate the pa-

[31] A school input that we cannot measure is the curriculum content within the classroom. A proxy for curriculum could be the grade level the child is currently attending. However, to the extent that grade progression depends on prior achievement, grade level would reflect all past inputs and would be inappropriate to include. If grade progression were automatic, then age effects included in our specification would capture grade-specific curriculum content.

[32] As described in the previous section, we allow the coefficients associated with lagged home inputs measured at ages 3–5 to differ from those at other ages, because the questions in the home scale were different at that age.

TABLE 3

ALTERNATIVE SPECIFICATIONS OF THE PRODUCTION FUNCTION

	CONTEMPORANEOUS	CUMULATIVE	VALUE-ADDED (1)	VALUE-ADDED (2)	SIBLING FIXED EFFECTS	CHILD FIXED EFFECTS
	A. Math Percentile Score*					
Home inputs:						
Home (0)	.17 (.02)	.08 (.02)	.57 (.01)	.05 (.02)	.06 (.02)	.04 (.02)
Home (−1):						
Ages 6–7	. . .	.18 (.02)	. . .	. . .	.05 (.03)	.07 (.03)
Ages 8–13	. . .	.06 (.02)	. . .	.02 (.02)	.04 (.02)	.02 (.02)
Home (−2):						
Ages 8–9	. . .	.21 (.03)	. . .	.11 (.02)	.07 (.03)	.09 (.03)
Ages 10–13	. . .	.07 (.02)	. . .	.03 (.02)	.04 (.02)	.03 (.02)
Home (−3):						
Ages 10–11	. . .	.11 (.03)	. . .	. . .†	−.01 (.03)	. . .
Ages 12–13	. . .	−.02 (.03)	. . .	. . .†	−.03 (.03)	. . .
Home (−4):						
Ages 12–13	. . .	.10 (.04)	. . .	. . .†	. . .	. . .
Lag test score	. . .	-	.57 (.02)	.55 (.02)	. . .	. . .
AFQT	.34 (.02)	-			. . .	. . .
AFQT × ages 6–7	. . .	.28 (.02)	. . .	. . .	−.12 (.03)	−.11 (.03)
AFQT × ages 8–9	. . .	.30 (.02)	.18 (.02)	.15 (.02)	−.07 (.03)	−.07 (.03)
AFQT × ages 10–11	. . .	.33 (.03)	.16 (.02)	.17 (.02)	−.03 (.03)	−.03 (.03)
AFQT × ages 12–13	. . .	.35 (.04)	.16 (.02)	.17 (.02)	. . .‡	. . .‡
Pupil-teacher ratio§	.01 (.03)	−.0.06 (.07)	−.007 (.02)	−.07 (.04)	−.02 (.08)	−.17 (.09)
Teacher salary	1.0E−4 (6.0E−5)	6.9E−5 (3.9E−5)	1.8E−5 (4.2E−5)	3.92E−5 (2.36E−5)	1.54E−5 (4.4E−5)	1.11E−5 (4.6E−5)
p-value for *F*-test of joint significance of school quality variables	.23	.20	.86	.19	.93	.11
	B. Reading Percentile Score*					
Home Inputs						
Home (0)	.17 (.02)	.11 (.02)	.10 (.02)	.06 (.02)	.09 (.02)	.03 (.02)
Home (−1):						
Ages 6–7	. . .	.13 (.03)	. . .	. . .	.01 (.03)	.01 (.03)
Ages 8–13	. . .	.05 (.02)	. . .	.02 (.02)	.03 (.02)	−.001 (.02)
Home (−2):						
Ages 8–9	. . .	.18 (.03)	. . .	.14 (.03)	.07 (.03)	.09 (.03)
Ages 10–13	. . .	.05 (.02)	. . .	.04 (.02)	.04 (.02)	.01 (.02)
Home (−3):						
Ages 10–11	. . .	.06 (.03)	. . .	. . .†	−.03 (.03)	. . .
Ages 12–13	. . .	−.01 (.03)	. . .	. . .†	−.02 (.03)	. . .

TABLE 3
(*Continued*)

	Contemporaneous	Cumulative	Value-Added (1)	Value-Added (2)	Sibling Fixed Effects	Child Fixed Effects
Home (−4):						
Ages 12–13	. . .	.09	. . .	. . .†	. . .	. . .
		(.04)				
Lag test score	. . .	. . .	.47	.47	. . .	. . .
			(.01)	(.02)		
AFQT	.29	. . .			. . .	. . .
	(.02)					
AFQT × ages 6–7	. . .	.21	. . .	. . .	−.15	−.13
		(.02)			(.03)	(.03)
AFQT × ages 8–9	. . .	.25	.21	.17	−.10	−.08
		(.02)	(.02)	(.02)	(.03)	(.03)
AFQT × ages 10–11	. . .	.33	.21	.21	−.03	−.01
		(.03)	(.02)	(.02)	(.03)	(.03)
AFQT × ages 12–13	. . .	.35	.22	.22	. . .‡	. . .‡
		(.03)	(.03)	(.03)		
Pupil-teacher ratio§	.04	−.06	.02	−.01	.02	.05
	(.03)	(.07)	(.02)	(.04)	(.09)	(.10)
Teacher salary	1.05E-5	7.31E-5	8.2E-5	4.5E-5	3.1E-6	5.7E-5
	(6.2E-5)	(3.7E-5)	(4.3E-5)	(2.3E-5)	(4.9E-5)	(5.2E-5)
p-value for *F*-test of joint significance of school quality variables	.11	.13	.09	.19	.97	.30

Note.—Robust standard errors are in parentheses.
* Also includes birth weight, first- and second-born dummies, dummies for mother's age at birth 18–19 and 20–29, child's age in months and its square, dummies for child age in years, mother's schooling, race and sex dummies, and cumulative family income.
† Omitted because of joint insignificance.
‡ In the fixed-effect specifications, one category has to be omitted. The coefficients on the other ages are interpretable as relative to the effect of the omitted category.
§ For the contemporaneous and value-added (1) specifications, the included school quality measures refer to contemporaneous school quality. For the other specifications, the school quality measures are cumulative as described in the text.

rameter estimates for the hybrid model. A comparison of the nonhybrid and hybrid specifications shows that the estimates are relatively insensitive to the inclusion of family income, race, and gender (included only in the hybrid model).

The fourth columns of tables 3 and A4 (labeled Value Added (2)) present the estimates for the value-added plus model, which adds to the strict value-added model the lagged input variables. Only two lagged inputs are included because the additional lags were not precisely estimated. For both the reading and math scores, the additional lagged inputs are individually and jointly statistically different from zero. Mother's AFQT continues to be a significant determinant of child test scores in the value-added model specifications, although its estimated effect is roughly constant and is not increasing in child age as was observed in the cumulative model.

The fifth and sixth columns of tables 3 and A4 present estimates for the within-family and within-child fixed-effect estimators. The contemporaneous home input and most of the lagged input variables continue to be statistically significant determinants of test scores in both the sibling fixed-effect and the child fixed-effect specifications. Some of the coefficients associated with lagged inputs are sensitive whether the within-family or the within-child estimator is used, but they are not very sensitive to whether the nonhybrid or hybrid specifications are used. For the fixed-effect specifications, we again see the pattern that the mother's AFQT effect increases in magnitude as children get older.[33]

In addition to the results reported in the tables, we also estimated the full set of specifications in which the pupil-teacher ratio was measured at the county level rather than at the state level.[34] The county-level measure was even less often statistically significant, which is consistent with findings reported in the literature. For example, Card and Krueger (1996*a*) found significant effects of state-level school quality measures on earnings. Betts (1995) compares the estimated effect of pupil-teacher ratios and teacher experience on earnings under different levels of aggregation of the quality measures and finds that the quality measures are more often significant when measured at the state level.[35]

The results reported in the tables do not take into account information on biological fathers, except to the extent that the home score may reflect in part activities of the father (e.g., buying books, taking children to museums, or involving them in lessons). The information on fathers is often missing in the data set, and requiring that it be present would necessitate dropping many observations. In preliminary analyses we found an indicator for whether the biological father was present in the household to not be statistically significant.

To summarize, the estimated models provide strong evidence that home inputs are important determinants of child test scores and that test scores depend on both the current and historical home environment as well as on mother's abilities as measured by the AFQT. We did not find evidence of a significant relationship between the schooling input measures and the test scores.[36]

[33] The left-out category is ages 12–13, so that the observed negative estimated coefficients are relative to the oldest age.

[34] Information on county of residence is the most detailed measure of location available for the NLSY79 respondents.

[35] Our state-level measure of quality differs in some ways from measures in the literature, where state often corresponds to a person's state of birth and it is assumed that the child is educated in his or her state of birth. In our case, the state measure gives the state-level average quality at the time of the child's residence. If a child moves from one state to another, our average school input measures would change to reflect different levels of school inputs across states and to reflect the amount of time spent in each location.

[36] As noted in the introduction, the lack of a relationship either could be a result of measurement error in the aggregated quality measures or may indicate that teacher salaries and pupil-teacher ratios do not adequately measure school quality, as argued in Hanushek (1986).

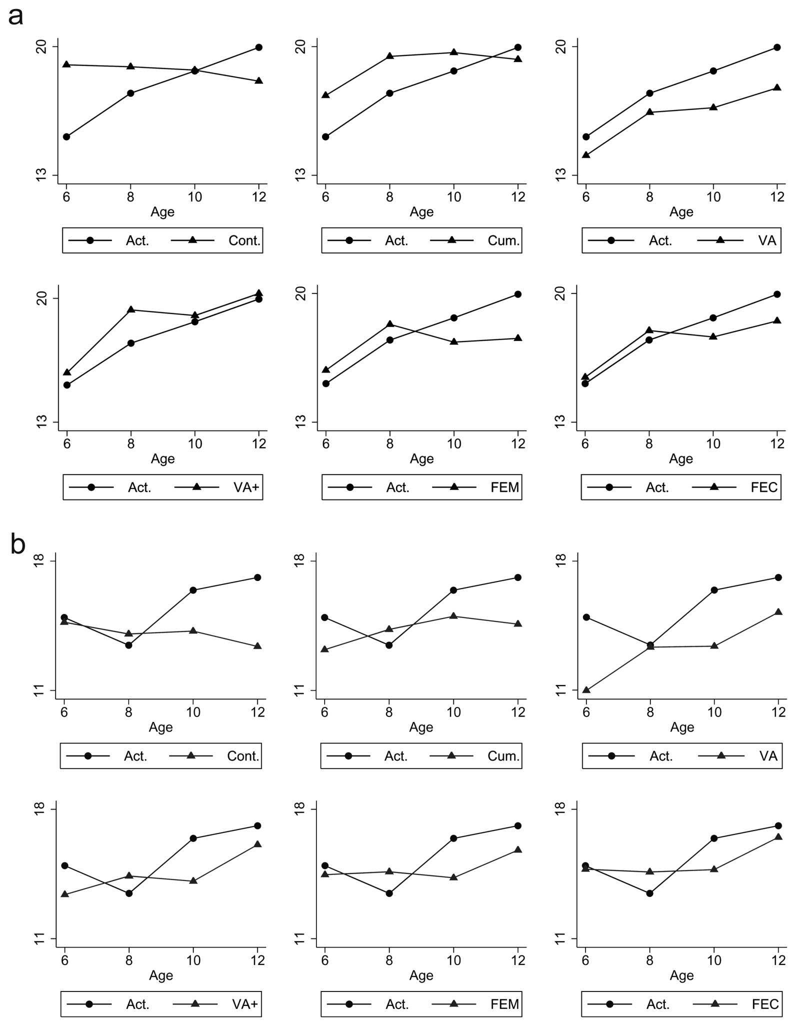


Figure 3.—*a*, Black-white difference in PIAT math percentage for girls, hybrid production function. *b*, Hispanic-white difference in PIAT math percentage for girls, hybrid production function.

C. *Within-Sample Goodness of Fit*

As seen in tables 3 and A4, inferences about the magnitude of the contribution inputs in explaining test scores depend to some extent on model specification. We next visually assess the performance of the alternative models in reproducing the test score gap patterns and how the gaps patterns vary by age, race/ethnicity, and gender. Figures 3*a*–*h* present the predicted gap pattern superimposed on the actual gap pat-

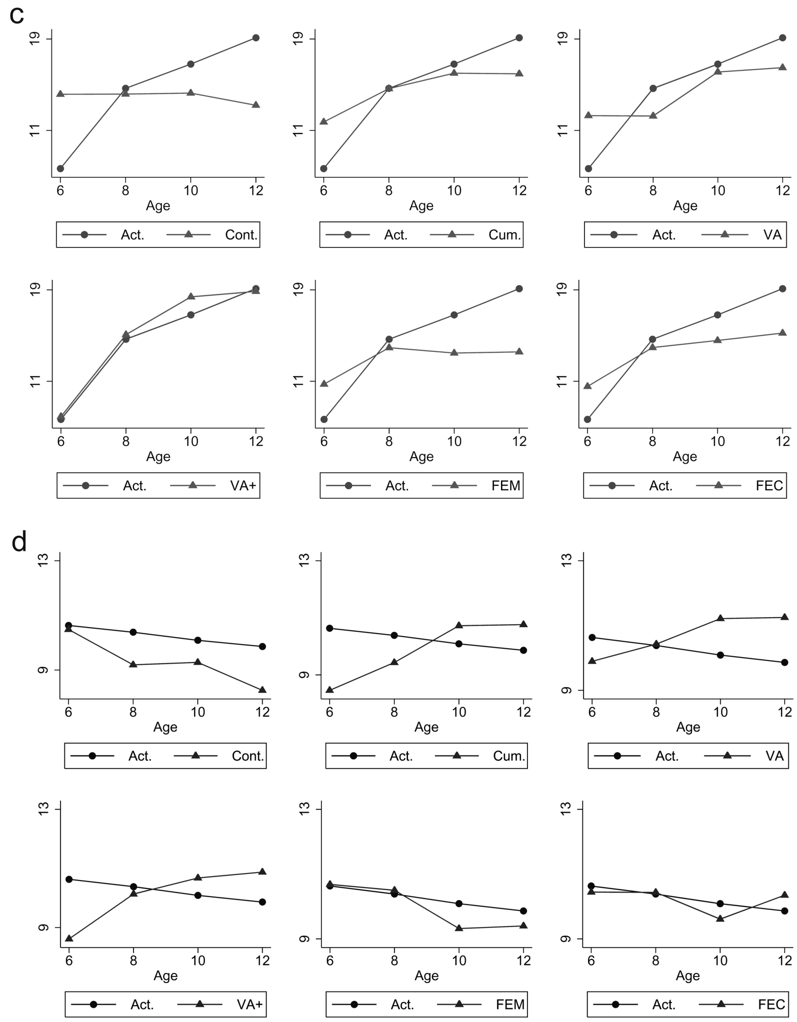


Figure 3.—*c*, Black-white difference in PIAT reading percentage for girls, hybrid production function. *d*, Hispanic-white difference in PIAT reading percentage for girls, hybrid production function.

tern for the alternative specifications for the hybrid versions of the production function and for the math and reading percentile scores, by race/ethnicity and gender. The hybrid model includes race, gender, and cumulative family income as additional covariates, which considerably improves the fit to the gap patterns. The nonhybrid model is not shown for the sake of brevity. The fixed-effect models (fixed-effect mother [FEM] and fixed-effect child [FEC]) can be viewed as variants of hybrid models because the fixed effects implicitly include race/eth-

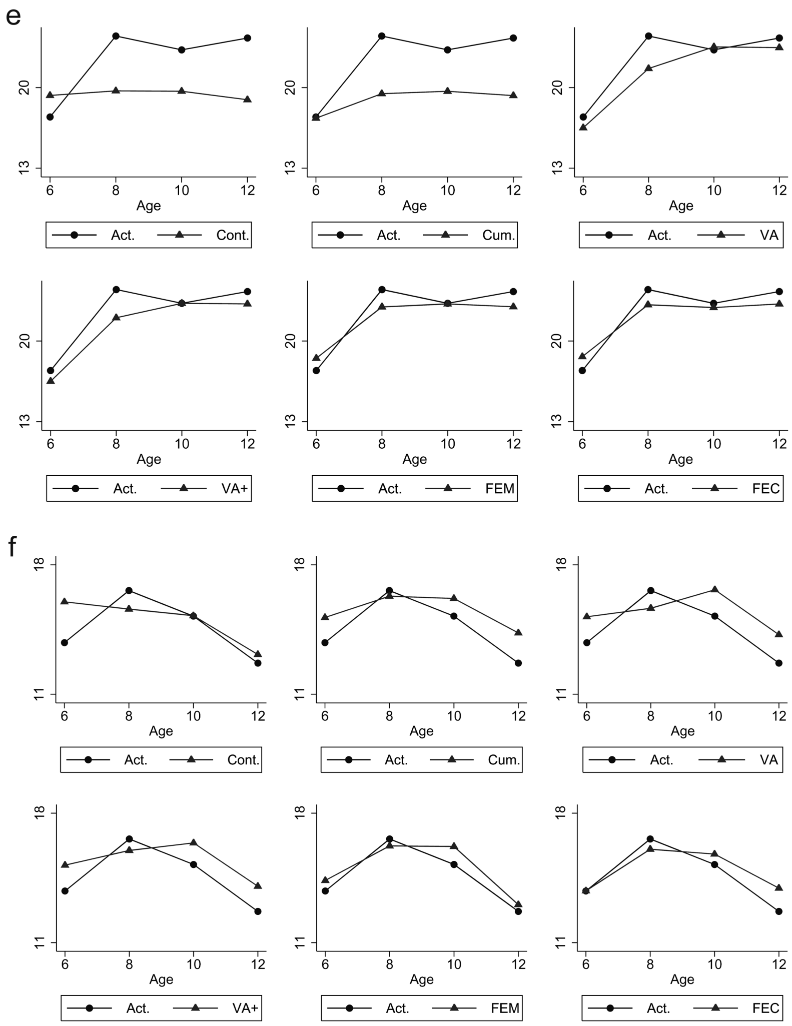


Figure 3.—*e*, Black-white difference in PIAT math percentage for boys, hybrid production function. *f*, Hispanic-white difference in PIAT math percentage for boys, hybrid production function.

nicity and gender. As seen in the figures, the contemporaneous specification usually does not reproduce the gap pattern, either for boys or for girls. In particular, it fails to generate the observed widening gap for both black girls and boys. The cumulative model without endowments better reproduces the rising gap pattern, but the fit is generally not as good as for the value-added (VA), value-added plus (VA+), and fixed-effect models. There are some subgroups for which the value-added plus model is clearly a better fit than the fixed-effects models.

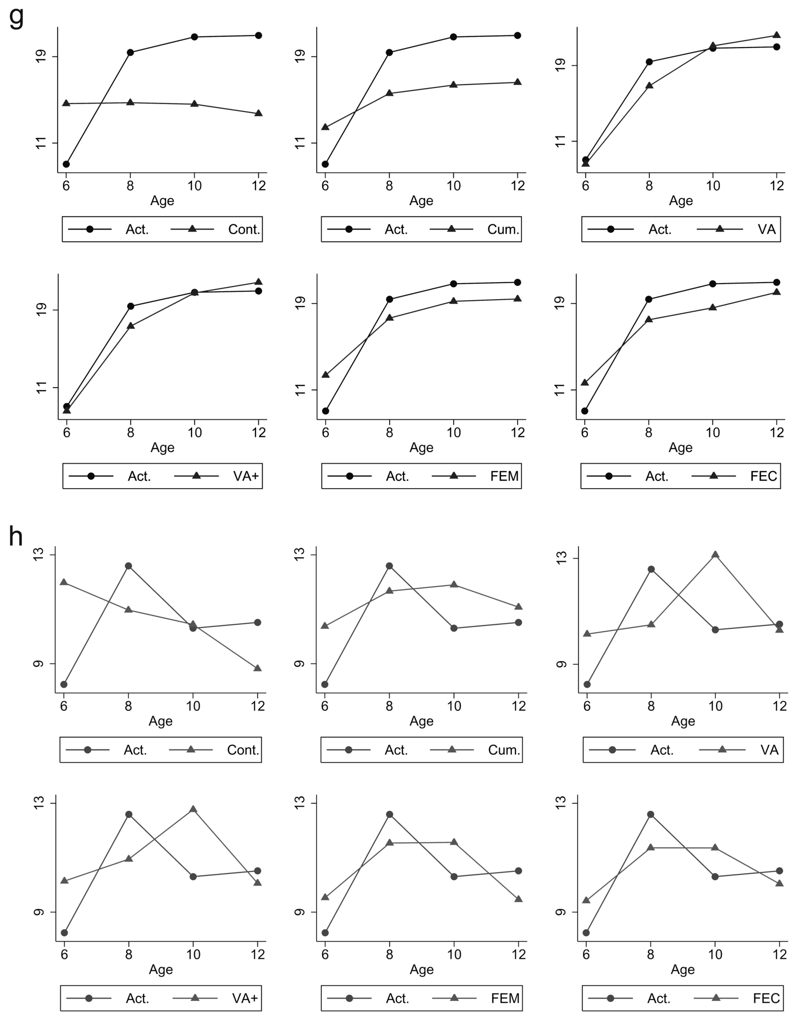


Figure 3—*g*, Black-white difference in PIAT reading percentage for boys, hybrid production function. *h*, Hispanic-white difference in PIAT reading percentage for boys, hybrid production function.

D. Model Selection

As noted in Section I, one of our goals in estimating the production function is to decompose the observed racial/ethnic test score gaps into components due to home, school, and mother's AFQT differences. When we evaluate the average test scores of the minority groups at the average home input levels observed for whites, we are in essence using the model to perform an out-of-sample forecast. We therefore use cross-validation methods to explore the reliability of the estimated models for out-of-sample forecasting purposes. These methods compare models on the basis of an out-of-sample RMSE criterion and can be used to compare nonnested models.[37]

Cross-validation approaches repeatedly estimate the model on a subset of the data, alternating which part of the sample is left out. We perform the cross-validation in two ways: using random holdout samples, which is the conventional approach, and using selective holdout samples that correspond to groups based on race and gender. The cross-validation procedure based on random holdout samples is implemented as follows. First, the entire sample is randomly divided into six roughly equal-sized subsamples. The model is repeatedly estimated on five of the six subsamples and used to construct the RMSE for the left-out subsample, alternating which group is left out. The RMSE values for each subsample are then summed to obtain the overall RMSE for that model. We constructed the overall RMSE for three different initial randomizations and report in table 4 the CV-RMSE for each of the production function specifications that was shown in table A4. Panel A of table 4 reports the cross-validation results for the nonhybrid specifications and panel B for the hybrid specifications, which include as additional covariates family income, gender, and race.

For the value-added specifications that include lagged test scores as covariates, there is a question of how to forecast lagged test scores in a way that does not make any use of data from the holdout sample. We first estimate the model for ages 6–7, at which there is no lagged test score. Then we use the model to forecast the ages 6–7 test scores, which are used as the lagged test score at ages 8–9. The ages 8–9 forecast becomes the lagged test score at ages 9–10 and so on.

Similarly, for the fixed-effect specifications, we first estimate the fixed-effect specifications excluding any data from the holdout samples. For the subsample used in estimation, we regress the estimated fixed effects on the observed covariates. The estimated model for fixed effects is used to generate predicted fixed effects for the holdout sample.

We also performed the cross-validation exercise for nonrandom holdout samples, leaving out one race × gender group at a time. As seen in table 4, under either the random holdout sample criterion or the selective holdout sample, the model specification that repeatedly tends to exhibit the lowest CV-RMSE is the value-added model with additional lagged inputs, which is the best six out of eight times.[38] Again, the value-

[37] For models estimated by maximum likelihood, the cross-validation criterion with random holdout samples (also called training samples) has been shown to be asymptotically equivalent to the Akaike model selection criterion (see Stone 1976). A good, nontechnical introduction to cross-validation can be found in Hastie, Tibshirani, and Friedman (2001).

[38] In some cases, the CV-RMSE values associated with two different specifications are quite close, differing in the third digit. In our experience, in applying cross-validation methods in other contexts, it is common to observe relatively small differences in RMSE.

TABLE 4
CROSS-VALIDATION RMSE FOR ALTERNATIVE SPECIFICATIONS: PERCENTILE AND RAW SCORES

	RANDOM HOLDOUT SAMPLE*				RACE × SEX HOLDOUT SAMPLE†			
	Math		Reading		Math		Reading	
	Percentile	Raw	Percentile	Raw	Percentile	Raw	Percentile	Raw
	A. Production Function with Baseline Variables							
Contemporaneous	23.46	8.31	25.67	8.51	24.38	8.60	26.39	8.71
Cumulative	23.20	8.20	25.50	8.38	24.09‡	8.49	26.23	8.58
Value-added	23.24	8.23	25.51	8.40	24.18	8.53	26.37	8.61
Value-added plus lags	23.20‡	8.20‡	25.50‡	8.39	23.99	8.45‡	26.21‡	8.56‡
Sibling fixed effects	23.20	8.20	25.50	8.38‡	24.31	8.56	26.72	8.68
Child fixed effects	31.01	10.99	25.65	8.51	34.62	11.61	32.18	9.36
	B. "Hybrid" Production Function							
Contemporaneous	23.29	8.25	25.55	8.46	24.35	8.59	26.39	8.70
Cumulative	23.09	8.16	25.39	8.35	24.08	8.49	26.22	8.57
Value-added	23.08	8.17	25.31	8.35	24.20	8.53	26.37	8.60
Value-added plus lags	23.06‡	8.15‡	25.30‡	8.34‡	24.00‡	8.45‡	26.20‡	8.55‡
Sibling fixed effects	23.09	8.17	25.41	8.35	24.33	8.59	26.85	8.71
Child fixed effects	28.35	10.15	25.57	8.45	35.51	12.93	31.46	9.31

* Based on six random holdout samples. Model is estimated on five of the six groups and used to generate RMSE for the left-out groups. The number shown is the average RMSE based on three replications of this procedure.
† Based on six race/sex groups. Model is estimated on one race × sex group (e.g., white boys) and used to generate the RMSE for other five race × sex groups.
‡ Denotes specification with the smallest RMSE value. In some cases, values differ in the third digit.

added model with additional lagged inputs emerges as the one with the lowest RMSE among the set of specifications considered.

V. Accounting for Sources of Racial Test Score Gaps

Using the production function estimates from the last section, we examine the extent to which differences in inputs and in mother's AFQT can account for racial/ethnic disparities in test scores. The implied impacts of our schooling quality measures on test scores were very small in comparison to those of the home inputs and AFQT, so we do not include them in the decomposition.[39] The decomposition results are presented for the value-added plus model, which was the preferred model according to the cross-validation criterion.

Panel A of table 5 examines how eliminating the gap in home inputs and in mother's AFQT would close the gap in test scores. That is, we assess the extent of the gap if black average home inputs were set to the level observed for white children and if black mothers' AFQT scores were set to the average level of white children.

The column labeled Actual Gap shows the percentile and raw score gap separately by gender and averaged over all ages (6–13) for both the reading and the math tests. The columns labeled Predicted Gap give the predicted gap according to the nonhybrid (1) and hybrid (2) specifications of the production function. The columns labeled Closed by Home give the amount of the predicted gap that would be closed if black children received, on average, the level of white home inputs, with all other inputs held constant at their observed levels. Roughly 10–20 percent of the test score gap in math and reading would be closed by equalizing home inputs. The columns labeled Closed by AFQT show the amount of the gap that is closed by equalizing mother's AFQT, again with all other inputs held constant. The estimates indicate that roughly half of the predicted gap is accounted for by differences in mother's AFQT. To the extent that higher AFQT mothers choose better home environments for the children, the partial effect of increasing mother's AFQT understates the total effect, which would be inclusive of increases in home inputs.

Panel B of table 5 presents analogous results for the white-Hispanic decomposition. The findings concerning the relative contribution of mother's AFQT and home inputs are generally similar to those of the black-white decomposition. One difference is that home inputs explain a larger proportion of the predicted reading gap for Hispanics than for blacks. To summarize, the decomposition estimates indicate that racial/ethnic differences in mother's AFQT generally explain a large fraction of the gap in test scores. However, the contribution of home inputs is not negligible. Our estimates thus imply that equalizing home inputs

[39] Given AFQT, mother's schooling also accounts for little of the gap.

TABLE 5
RACIAL GAP CLOSED BY HOME INPUTS AND BY MOTHER'S AFQT
SPECIFICATION: VALUE-ADDED MODEL WITH LAGS

	Actual Gap	Predicted Gap* (1)	Predicted Gap* (2)	Closed by Home (1)	Closed by Home (2)	Closed by AFQT (1)	Closed by AFQT (2)
				A. Black-White Gap			
Math percentile:							
Boys	22.37	14.96	19.41	2.64 (11.8%)	2.30 (10.3%)	10.69 (47.8%)	8.31 (37.1%)
Girls	17.80	14.18	15.01	2.51 (14.1%)	2.32 (13.0%)	10.42 (58.5%)	6.42 (36.1%)
Reading percentile:							
Boys	17.56	13.84	14.43	2.75 (15.7%)	2.55 (14.5%)	9.22 (52.5%)	8.88 (50.6%)
Girls	14.56	13.45	13.95	2.62 (18.0%)	2.43 (16.7%)	8.99 (61.7%)	8.66 (59.4%)
Math raw score:							
Boys	7.80	5.13	6.60	.89 (11.4%)	.77 (9.9%)	3.73 (47.8%)	2.92 (37.4%)
Girls	6.07	4.83	6.31	.85 (14.0%)	.74 (12.2%)	3.63 (59.8%)	2.84 (46.8%)
Reading raw score:							
Boys	6.78	4.87	5.31	.88 (13.0%)	.80 (11.8%)	3.38 (49.9%)	3.15 (46.4%)
Girls	5.36	4.70	5.11	.84 (12.4%)	.76 (11.2%)	3.30 (48.7%)	3.07 (45.3%)
				B. Hispanic-White Gap			
Math percentile:							
Boys	14.58	11.91	15.00	2.66 (18.2%)	2.32 (15.9%)	8.26 (56.7%)	6.41 (44.0%)
Girls	15.48	11.18	14.09	1.73 (11.2%)	1.51 (9.8%)	8.33 (53.8%)	6.47 (41.8%)
Reading percentile:							
Boys	10.41	11.19	10.55	2.78 (26.7%)	2.58 (24.8%)	7.12 (68.4%)	6.86 (65.9%)
Girls	10.24	10.17	9.30	1.81 (17.7%)	1.68 (16.4%)	7.18 (70.1%)	6.92 (67.6%)
Math raw score:							
Boys	5.47	4.47	5.45	.90 (16.4%)	.78 (14.3%)	2.88 (52.7%)	2.25 (41.1%)
Girls	5.47	3.97	4.85	.59 (10.8%)	.51 (9.3%)	2.90 (53%)	2.27 (41.4%)
Reading raw score:							
Boys	4.30	4.30	3.99	.89 (20.7%)	.80 (18.6%)	2.61 (60.7%)	2.43 (56.5%)
Girls	3.57	3.68	3.26	.58 (16.2%)	.52 (14.6%)	2.63 (73.7%)	2.45 (68.6%)

Note.—The percentage of the gap closed is in parentheses.
* (1) is the nonhybrid model and (2) is the hybrid model.

of whites and blacks, holding all other inputs constant, would close a significant proportion (10–20 percent) of the test score gap. Also, increasing the home environment of the current generation would presumably increase their AFQT scores as adults, which would also have benefits for future generations.

VI. Conclusions

This article considered ways of estimating the cognitive achievement production function that are consistent with theoretical notions that achievement is a cumulative process depending potentially on the entire history of family and school inputs as well as on parental abilities and unobserved endowments. Using rich longitudinal data, we implemented alternative specifications of the production function. Across almost all the specifications considered, we found that mother's accumulated ability, as measured by the AFQT, and home inputs (contemporaneous and lagged) are substantive determinants of children's test scores in math and reading. The estimated magnitude of lagged home input effects is often similar to that of current inputs, and the effect of current inputs tends to be overstated in specifications that ignore lagged input effects. The coefficients associated with the school inputs were for the most part not precisely estimated.

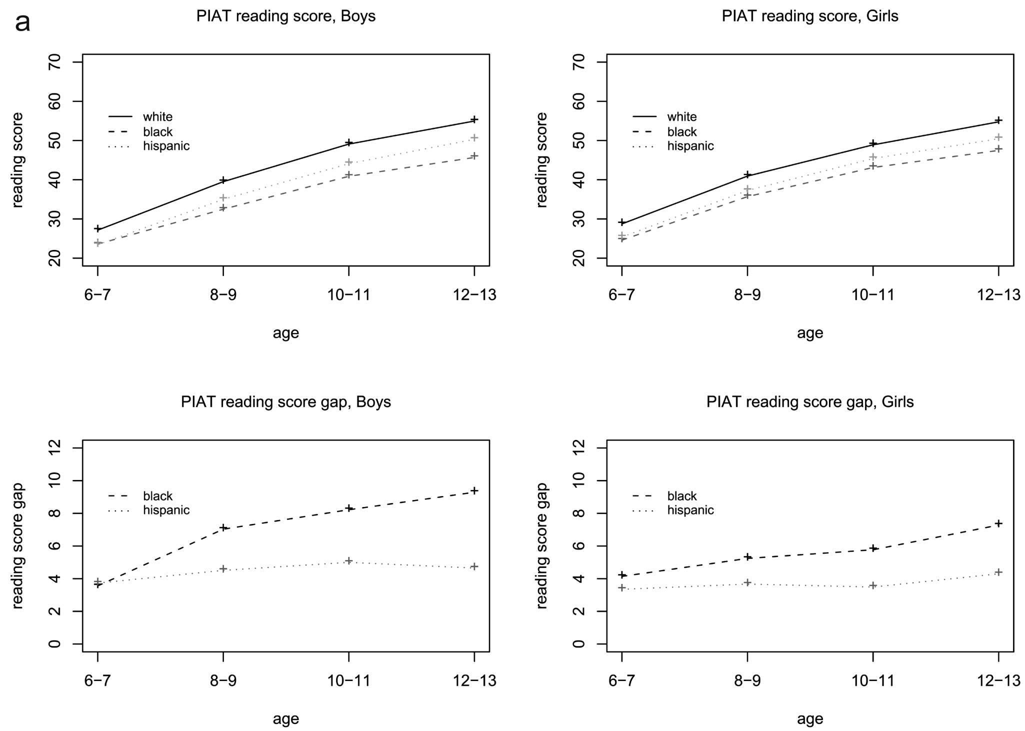
When alternative model specifications of the production function are compared using a cross-validation criterion, we found the most support for the value-added plus model, which augments a basic value-added model with additional lagged input variables. We used the production function parameter estimates for this preferred specification to examine the sources of racial/ethnic test score gaps. Differences in mother's AFQT test scores account for the largest portion of the black-white and Hispanic-white test score gaps—roughly half for both reading and math. Differences in home inputs account for 10–20 percent of the black-white and the Hispanic-white test score gaps. Differences in school inputs and in mother's schooling account for only very small portions of the gap.

Our findings do not have direct policy implications for the most efficient way to reduce the test score gap. What is required to make such a determination is knowledge of the relative costs of alternative policies and of how schools and parents make input decisions, to account for the possibility that changing the level of a single input affects decisions about other inputs. A full assessment of such policies would require a more complete analysis of how families make decisions about what inputs to provide for their children.[40]

[40] For example, see recent efforts by Mroz, Liu, and van der Klaauw (2003) and Bernal and Keane (2005).

Appendix

a


PIAT reading score, Boys
PIAT reading score, Girls
PIAT reading score gap, Boys
PIAT reading score gap, Girls
reading score
reading score gap
age
white
black
hispanic
6−7
8−9
10−11
12−13
20
30
40
50
60
70
0
2
4
6
8
10
12

b

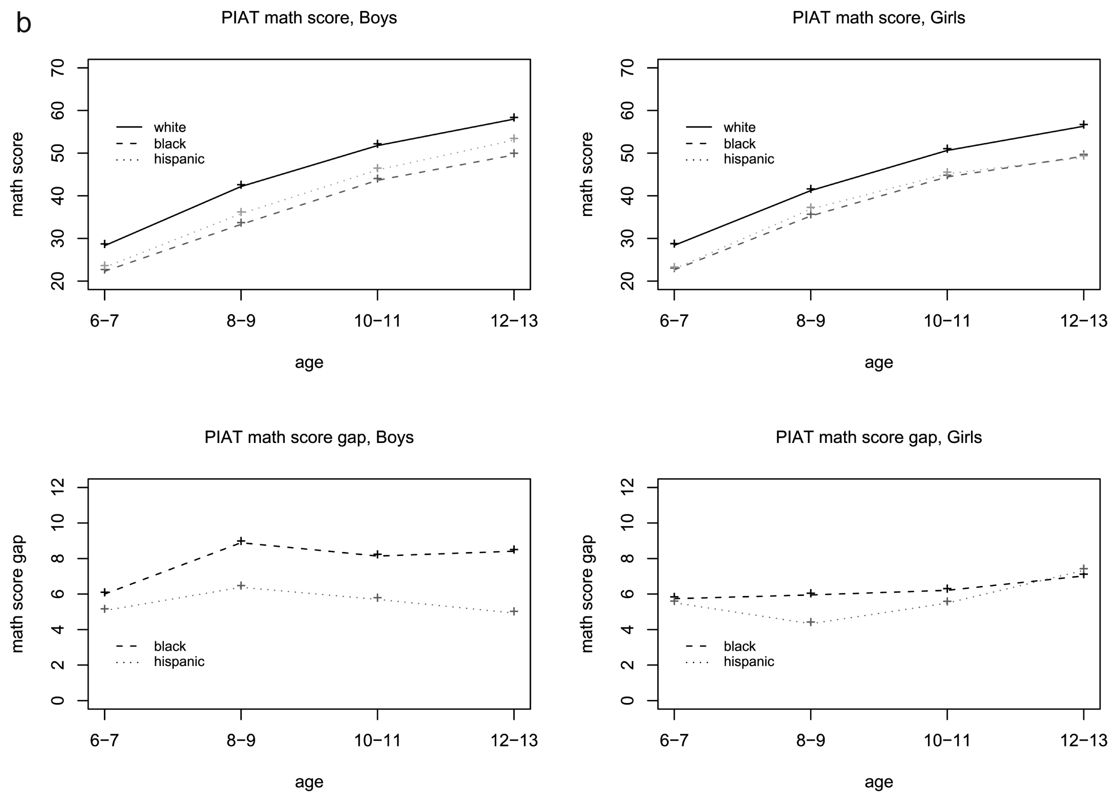


Figure A1.—Comparison of raw scores by age by race/ethnicity: a, reading; b, math

TABLE A1
COMPARISON OF RESPONSES ON INDIVIDUAL HOME INPUT SCORE ITEMS BY RACE: AGE 3–5 QUESTIONS

Question	White	African American	Hispanic
	Asked of Mother		
How many children's books does your child have? (1 if > 9 books)	94	57	63
How often do you read stories to your child? (1 if at least 3 times a week)	70	40	44
How many magazines does your family get regularly? (1 if at least 3)	41	31	26
Does your child have the use of a compact disk player or tape recorder or record player and at least 5 children's records or tapes?	83	66	67
Do you help your child learn numbers?	78	63	70
Do you help your child learn the alphabet?	77	63	66
Do you help your child learn colors	78	62	70
Do you help your child learn shapes/sizes?	72	51	56
How often does a family member take the child on any kind of outing?	87	74	77
How often did family member arrange visit to museum within last year?	72	66	61
	Interviewer Observations		
Child's play environment is safe (no dangerous health or structural hazards within a preschooler's range)	95	90	91
Is interior of the home dark or perceptually monotonous?	5	16	8
Are visible rooms reasonably clean?	94	90	92
Are visible rooms minimally cluttered?	84	82	84

Note.—Numbers denote the percentage receiving the highest score, where the highest score is 1 and lowest score is 0.

TABLE A2
Comparison of Responses on Individual Home Input Score Items by Race: Age 6–9 Questions

Question	White	African American	Hispanic
	Asked of Mother		
How many books does your child have? (1 if > 9 books)	94	64	68
How often do you read to your child?	45	28	32
Is there a musical instrument that your child can use at home?	47	30	33
Does family get a daily newspaper?	51	41	40
How often does child read for enjoyment?	74	72	68
Does family encourage child to start and keep hobbies?	93	86	83
Does child get special lessons or belong to organizations that encourage activities (sports, arts)	61	41	39
How often did family member arrange visit to museum within last year?	80	70	70
How often did family member take child to musical or theatrical performance within the past year?	61	56	49
When family watches TV together, do you or child's father discuss TV programs with him/her?	88	71	79
	Interviewer Observations		
Is interior of the home dark or perceptually monotonous?	5	15	7
Are visible rooms reasonably clean?	93	90	92
Are visible rooms minimally cluttered?	84	83	84
Building has no potentially dangerous structural or health hazards withins a school-aged child's range	71	65	70

Note.—Numbers denote the percentage receiving the highest score, where the highest score is 1 and the lowest score is 0.

TABLE A3

Comparison of Responses on Individual Home Input Score Items by Race: Age 10–13 Questions

Question	White	African American	Hispanic
	Asked of Mother		
How many books does your child have? (1 if > 19 books)	75	38	43
Is there a musical instrument that your child can use at home?	55	32	35
Does family get a daily newspaper?	49	41	41
How often does child read for enjoyment?	66	64	65
Does family encourage child to start and keep hobbies?	95	89	88
Does child get special lessons or belong to organizations that encourage activities (sports, arts)?	69	56	50
How often did family member arrange visit to museum within last year?	80	72	69
How often did family member take child to musical or theatrical performance within the past year?	61	59	50
When family watches TV together, do you or child's father discuss TV programs with him/her?	87	69	74
	Interviewer Observations		
Is interior of the home dark or perceptually monotonous?	5	15	8
Are visible rooms reasonably clean?	93	91	93
Are visible rooms minimally cluttered?	85	83	83
Building has no potentially dangerous structural or health hazards within a school-aged child's range	70	65	66

Note.—Numbers denote the percentage receiving the highest score, where the highest score is 1 and the lowest score is 0.

TABLE A4

Alternative Specifications of the "Hybrid" Production Function

	Contemporaneous	Cumulative	Value Added (1)	Value Added (2)	Sibling Fixed Effects	Child Fixed Effects
	A. Math Percentile Score*					
Home inputs:						
Home (0)	.15	.07	.08	.05	.06	.04
	(.02)	(.02)	(.01)	(.02)	(.02)	(.02)
Home (−1):						
Ages 6–7	. . .	.17	. . .	. . .	.05	.07
		(.02)			(.03)	(.03)
Ages 8–13	. . .	.05	. . .	.02	.04	.02
		(.02)		(.02)	(.02)	(.02)
Home (−2):						
Ages 8–9	. . .	.19	. . .	.10	.07	.10
		(.03)		(.02)	(.03)	(.03)
Ages 10–13	. . .	.06	. . .	.02	.04	.03
		(.02)		(.02)	(.02)	(.02)
Home (−3):						
Ages 10–11	. . .	.10	. . .	. . .[†]	−.01	. . .
		(.03)			(.03)	
Ages 12–13	. . .	−.03	. . .	. . .[†]	−.03	. . .
		(.03)			(.03)	
Home (−4):		.09				
Ages 12–13	. . .	(.04)	. . .	. . .[†]	. . .	. . .
Lag test score	. . .	. . .	.56	.55	. . .	. . .
			(.02)	(.02)		
AFQT	.26	.27	. . .		. . .	. . .
	(.02)	(.03)				
AFQT × ages 6–7	. . .	−.06	. . .	. . .	−.12	−.10
		(.03)			(.03)	(.03)
AFQT × ages 8–9	. . .	−.03	.14	.12	−.07	−.06
		(.03)	(.02)	(.02)	(.03)	(.03)
AFQT × ages 10–11	. . .	. . .	.12	.13	−.03	−.03
			(.03)	(.03)	(.03)	(.03)
AFQT × ages 12–13	. . .	.02	.13	.14	. . .[‡]	. . .[‡]
		(.03)	(.02)	(.03)		
Pupil-teacher ratio	.01	−.06	−.01	−.07	−.02	−.18
	(.03)	(.07)	(.02)	(.05)	(.08)	(.09)
Teacher salary	7.36E−5	5.6E−5	5.86E−7	3.2E5	1.6E5	7.7E6
	(5.99E−5)	(3.8E−5)	(4.16E−5)	(2.3E5)	(4.4E5)	(4.7E5)
p-value for *F*-test of joint significance of school quality variables	.44	.33	.91	.22	.93	.10
	B. Reading Percentile Score*					
Home inputs:						
Home (0)	.16	.10	.09	.05	.08	.03
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Home (−1):						
Ages 6–7	. . .	.13	. . .	. . .	.01	.01
		(.03)			(.03)	(.03)
Ages 8–13	. . .	.04	. . .	.02	.02	−.003
		(.02)		(.02)	(.02)	(.02)
Home (−2):						
Ages 8–9	. . .	.18	. . .	.13	.06	.09
		(.03)		(.03)	(.03)	(.03)
Ages 10–13	. . .	.05	. . .	.04	.03	.01
		(.02)		(.02)	(.02)	(.02)

TABLE A4
(*Continued*)

	Contempo-raneous	Cumulative	Value Added (1)	Value Added (2)	Sibling Fixed Effects	Child Fixed Effects
Home (−3):						
Ages 10–11	. . .	.06	. . .	. . .†	−.03	. . .
		(.03)			(.03)	
Ages 12–13	. . .	-0.01	. . .	. . .†	−.02	. . .
		(.03)			(.03)	
Home (−4):		.09				
Ages 12–13	. . .	(.04)	. . .	. . .†	. . .	. . .
Lag test score	. . .	. . .	.47	.47	. . .	. . .
			(.02)	(.01)		
AFQT	.28	. . .	…		. . .	. . .
	(.02)					
AFQT × ages 6–7	. . .	.21	. . .	. . .	−.15	−.12
		(.03)			(.03)	(.03)
AFQT × ages 8–9	. . .	.25	.17	.14	−.09	−.07
		(.03)	(.02)	(.02)	(.03)	(.03)
AFQT × ages 10–11	. . .	.33	.17	.18	−.02	−.01
		(.03)	(.02)	(.03)	(.04)	(.03)
AFQT × ages 12–13	. . .	.35	.18	.19	. . .‡	. . .‡
		(.04)	(.03)	(.03)		
Pupil-teacher ratio	.04	−.07	.02	−.03	.04	.04
	(.03)	(.07)	(.02)	(.04)	(.09)	(.10)
Teacher salary	9.3E−5	7.1E−5	5.8E5	3.8E5	−4.1E6	5.1E5
	(6.4E−5)	(3.7E−5)	(4.3E5)	(2.3E5)	(4.9E5)	(5.2E5)
p-value for *F*-test of joint significance of school quality variables	.19	.16	.28	.27	.90	.41

Note.—Robust standard errors are in parentheses.
* Also includes birth weight, first- and second-born dummies, dummies for mother's age at birth 18–19 and 20–29, child's age in months and its square, dummies for child age in years, mother's schooling, race and sex dummies, and cumulative family income.
† Omitted because of joint insignificance.
‡ In the fixed-effect specifications, one category has to be omitted. The coefficients on the other ages are interpretable as relative to the effect of the omitted category.

References

Altonji, J. G., and T. A. Dunn. 1996. "Using Siblings to Estimate the Effect of School Quality on Wages." *Rev. Econ. and Statis.* 78 (November): 665–71.

Ben-Porath, Y. 1967. "The Production of Human Capital and the Life-Cycle of Earnings." *J.P.E.* 75, no. 4, pt. 1 (August): 352–65.

Bernal, Raquel, and Michael Keane. 2005. "Maternal Time, Child Care and Child Cognitive Development: The Case of Single Mothers." Working paper, Northwestern Univ. and Yale Univ.

Betts, Julian. 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *Rev. Econ. and Statis.* 77 (May): 231–49.

Boardman, Anthony E., and R. J. Murnane. 1979. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." *Sociology Educ.* 52 (April): 113–21.

Boozer, Michael A., Alan B. Krueger, and Shari Wolken. 1992. "Race and School

Quality since Brown v. Board of Education." *Brookings Papers Econ. Activity: Microeconomics*, pp. 269–326.
Caldwell, R. H., and B. M. Bradley. 1984. *Home Observation for Measurement of the Environment.* Little Rock: Univ. Arkansas, Center Child Development and Educ.
Cameron, Steven, and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts." *J.P.E.* 106 (2): 262–333.
Carniero, Pedro, James J. Heckman, and Dimitriy V. Masterov. 2002. "Labor Market Discrimination and Racial Differences in Premarket Factors." Manuscript, Univ. Chicago.
Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Q.J.E.* 107 (February): 151–200.
———. 1996*a*. "Labor Market Effects of School Quality: Theory and Evidence." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Gary Burtless. Washington, DC: Brookings Inst.
———. 1996*b*. "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina." *J. Econ. Perspectives* 10 (4): 31–50.
Cook, Michael D., and William N. Evans. 2000. "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *J. Labor Econ.* 18 (4): 729–53.
Cunha, Flavio, and James J. Heckman. 2003. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." Working paper, Univ. Chicago.
Cunha, Flavio, James J. Heckman, Lance Lochner, and Dmitri Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, edited by E. A. Hanushek and F. Welch. Amsterdam: North-Holland.
Currie, Janet, and Duncan Thomas. 1999. "Early Test Scores, Socioeconomic Status, and Future Outcomes." Working Paper no. 6943, NBER, Cambridge, MA.
Donohue, J., J. J. Heckman, and P. E. Todd. 2002. "The Schooling of Southern Blacks: The Roles of Legal Activism and Private Philanthropy, 1910–1960." *Q.J.E.* 117 (February): 225–68.
Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Rev. Econ. and Statis.* 86 (May): 447–64.
———. 2005. "The Black-White Test Score Gap through the Third Grade." Working Paper no. 11049, NBER, Cambridge, MA.
Fuchs, V. R., and D. M. Reklis. 1994. "Mathematical Achievement in Eighth Grade: Interstate and Racial Differences" Working Paper no. 4284, NBER, Cambridge, MA.
Hanushek, E. A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *J. Econ. Literature* 24 (3): 1141–77.
Hanushek, E. A., and S. G. Rivkin. 2006. "School Quality and the Black-White Achievement Gap." Working Paper no. 12651, NBER, Cambridge, MA.
Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* New York: Springer-Verlag.
Hedges, L., R. D. Laine, and R. Greenwald. 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23 (3): 5–14.
Hedges, L., and A. Nowell. 1998. "The Black-White Test Score Convergence since 1965." In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Inst.

———. 1999. “Changes in the Black-White Gap in Achievement Test Scores.” *Sociology Educ.* 72 (April): 111–15.

Jencks, Christopher, and Meredith Phillips. 1998. “The Black-White Test Score Gap: An Introduction.” In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Inst.

Keane, M. P., and K. I. Wolpin. 1997. “Career Decisions of Young Men.” *J.P.E.* 105 (June): 473–522.

Krueger, Alan B. 1998. “Reassessing the View That American Schools Are Broken.” *Econ. Policy Rev.* 4 (March): 29–46.

———. 2003. “Economic Considerations and Class Size.” *Econ. J.* 113 (485): F34–F63.

Leibowitz, Arleen. 1974. “Home Investments in Children.” *J.P.E.* 82, no. 2, pt. 2 (March/April): S111–S131.

Ludwig, Jens. 2003. “The Great Unknown: Does the Black-White Test Score Gap Narrow or Widen through the School Years? It Depends on How You Measure.” *Education Next* 3 (Summer): 79–82.

Mroz, T., H. Liu, and W. van der Klaauw. 2003. “Maternal Employment, Migration and Child Development.” Working paper, Univ. North Carolina, Chapel Hill.

Murnane, R. J., R. A. Maynard, and J. C. Ohls. 1981. “Home Resources and Children’s Achievement.” *Rev. Econ. and Statis.* 63 (3): 369–77.

Murnane, R. J., J. B. Willett, K. L. Bub, and K. McCartney. 2006. “Understanding Trends in the Black-White Achievement Gaps during the First Years of School.” *Brookings-Wharton Papers Urban Affairs*, pp. 97–135.

Murnane, R. J., J. B. Willett, and F. Levy. 1995. “The Growing Importance of Cognitive Skills in Wage Determination.” *Rev. Econ. and Statis.* 77 (2): 251–66.

National Center for Education Statistics. 2005. *NAEP 2004: Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics.* Washington, DC: U.S. Dept. Educ.

Neal, D., and W. Johnson. 1996. “The Role of Pre-market Factors in Black-White Wage Differences.” *J.P.E.* 104 (October): 869–95.

Philips, M., J. Crouse, and J. Ralph. 1998. “Does the Black-White Test Score Gap Widen after Children Enter School?” In *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Inst.

Reardon, Sean. 2007. “Thirteen Ways of Looking at the Black-White Test Score Gap.” Manuscript, Stanford Univ.

Robertson, Donald, and James Symons. 1990. “The Occupational Choice of British Children.” *Econ. J.* 100 (402): 828–41.

Rosenzweig, M. 1986. “Birth Spacing and Sibling Inequality: Asymmetric Information within the Family.” *Indus. Econ. Rev.* 26 (3): 55–76.

Rosenzweig, M., and T. P. Schultz. 1982. “Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in Rural India.” *A.E.R.* 72 (September): 803–15.

Rosenzweig, M., and K. I. Wolpin. 1994. “Are There Increasing Returns to the Intergenerational Production of Human Capital? Maternal Schooling and Child Intellectual Achievement.” *J. Human Resources* 29 (Spring): 670–93.

———. 1995. “Sisters, Siblings, and Mothers: The Effects of Teen-age Childbearing on Birth Outcomes.” *Econometrica* 63 (March): 303–26.

Stone, M. 1976. “An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike’s Criterion.” *J. Royal Statis. Soc.*, ser. B, 39 (1): 44–47.

Tamura, Robert. 2001. “Teachers, Growth, and Convergence.” *J.P.E.* 109 (5): 1021–59.

Todd, Petra, and Kenneth I. Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *Econ. J.* 113 (485): F3–F33.