HCEO WORKING PAPER SERIES

Home and School in the Development of Children

Francesco Agostinelli
Morteza Saharkhiz
Matthew Wiswall

Working Paper

2019-044

07/2019

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Home and School in the Development of Children

Francesco Agostinelli
University of Pennsylvania

Morteza Saharkhiz
IMPAQ International LLC

Matthew Wiswall *
University of Wisconsin-Madison
NBER

June 2019

JEL Classification I21, J13
Keywords: Child Development; School; Parents; Education Production Function; Child Development; Skill Formation; Value-added; Latent Factor Models; Skill Production Technology

*This research is partially supported by NSF Grant 1357686. We thank participants at seminars at Arizona State University and University of Wisconsin-Madison for helpful comments. We are responsible for all errors.

Abstract

We develop a unified empirical framework for child development which nests the key features of two previously parallel research programs, the Child Development literature and the Education Production Function literature. Our framework allows for mis-measured cognitive and non-cognitive skills, classroom effects, parental influences, and complementarities. Although both are important, we estimate that differential parental investments are the more important source of end-of-kindergarten inequality than classroom quality. Quality classrooms have a larger effect on children entering kindergarten with skill deficits. Our estimated model replicates out-of-sample patterns by excluded race and family income variables and experimental results from the Tennessee STAR experiment.

1 Introduction

The wide dispersion of measured human capital in children and its strong relationship with later life outcomes has prompted a renewed interest in understanding the determinants of skill formation among children (for a recent review, see Heckman and Mosso, 2014). This paper develops a framework to analyze the key determinants of child developments during an important transition for many American children: entering formal schooling at kindergarten. For many children, this stage of their development represents a substantial increase in interactions with caregivers other than family. We analyze how the two crucial environments children face during childhood –home and school– affect the formation of children's skills.

Our paper unifies two largely separate and parallel research programs. First, research in the Child Development literature (such as work by Cunha and Heckman, 2007; Cunha et al., 2010; Agostinelli and Wiswall, 2016a; Attanasio et al., 2019a,b) uses survey data on measures of children's cognitive and non-cognitive skills to assess the importance of parental investments on the development process. This analysis is largely silent on the role of schools, and how the heterogeneity of school quality across children affects child development. In contrast, the Education Production Function literature, much of it using large scale administrative data from particular school systems (such as work by Rivkin et al., 2005; Krueger, 1999; Chetty et al., 2014a,b), focuses on estimating the value-added of classrooms and teachers. This research is largely silent on the role of influences outside school, and how the heterogeneity in home life affects development.

The starting point for our research is to develop a unified empirical framework that nests the key features of the prior research. We start by incorporating influences from both the home and school environments, treating both as latent, fundamentally unobserved, inputs with no perfect measure. For the home influences, we follow the latent factor structure of Cunha et al. (2010) by using a number of proxies for parental investments. For the latent school influences, we follow the education literature, and treat the school influences, at the classroom level, as a latent fixed effect, which is identified using data on multiple children in the same classroom.

Our model of child development follows Cunha et al. (2010) in allowing for both cognitive and non-cognitive skills, and follows the education literature in separating cognitive skills into mathematics and reading skills.[1] We model skill development in

[1]There is relatively little work using administrative school data to track non-cognitive skill development. One recent exception is Jackson (2012), which estimates teacher value-added models using North Carolina administrative data and absences and suspensions for 9th grade students as a proxy for non-cognitive outcomes.

each of these domains via a technology of skill formation, which defines the dynamics of children's skills through kindergarten. Each of the three types of skills is produced by the child's skills at kindergarten entry, and the inputs children are exposed to during kindergarten from their home and school/classroom environments. We follow the child development literature in allowing for complementarities between existing skills and current investments, implying that home and school environments can have heterogeneous effects on children.

Identification of this model requires solving two key challenges. The first challenge comes from the fact that we do not perfectly observe children's skills and parental investments in the data. Instead, we observe arbitrarily scaled measures which include measurement error. Previous research has shown that the estimates of the skill formation technology can suffer substantial bias if this issue is ignored (see Cunha et al., 2010; Agostinelli and Wiswall, 2016a). For this reason, we implement a latent factor model to recover the empirical distribution of math, reading, and non-cognitive latent skills for children from multiple test scores and measures, and use a similar model for the distribution of latent parental investments.

The second identification challenge comes from the selection bias: children may be systematically exposed to classrooms and parental investments based on their unobservable characteristics. This type of selection on unobservables would jeopardize our attempt to identify the school and home impacts on child development. We approach this issue in two ways. First, because of the rich nature of our data and the generality of our model, our identifying assumption is based on the exogeneity of a particular environment, once we control for a child's latent cognitive and non-cognitive skills and other features of the environment. This is a weaker assumption relative to either of the previous strands of the research: we use our data with measures of prior skills, home, and school investments to "fill-in" the variables that would otherwise be unobserved.

Second, we test our models for any remaining bias using a set of out-of-sample validation tests. One set of tests shows that our estimated model can replicate patterns of end-of-kindergarten skills across parental income and racial groups, even though these are excluded variables from our model. As in Chetty et al. (2014a), this suggests that are models are approximately "sufficient," and remaining bias is minimal. In a second set of validation exercises we replicate the experimental results of the Tennessee STAR experiment, in which children are randomly assigned kindergarten classrooms. Although we face the challenge of matching across datasets with different skill measures, the reasonably close match of our model predictions for this experiment to the actual observed experimental results gives us some confidence in our model estimates.

Because of the generality of our model, we face another practical challenge in computing the estimates for our model. The inclusion of non-linear classroom fixed effects and the interactions between latent classroom components with unobserved child skills (included to capture complementarities in skill formation) implies that our estimator does not have a simple closed form. To estimate our model, we therefore develop a multi-step iterative procedure, which builds upon previous work by Arcidiacono et al. (2012) to incorporate an instrumental variable estimator at each step. This procedure is tractable and transparent, and could be used in a number of contexts.

We find that both home and classroom investments are important inputs to child development during kindergarten. A 1 standard deviation increase in classroom quality has a 0.323, 0.381, and 0.519 impact on end-of-kindergarten mathematics, reading, and non-cognitive skills, respectively, after accounting for other inputs and measurement error. In contrast, the effects of a 1 standard deviation change in home environment has effect sizes that are between 15% and 40% lower than the classroom estimates, suggesting the greater importance of classroom. However, as we show in a series of counterfactual decomposition exercises, the quality of home environments are more closely related to household income than classroom environment (although of course high income households have higher quality in both), and therefore re-mediating income gaps in home environments has a larger impact on skill development than re-mediating gaps in classroom environments. Focusing on the gaps in skills between the 90th and 10th percentiles of the household income distribution, providing all children the classroom quality of what the high income, 90th percentile households receive, would decrease the skill gap between the 10th and the 90th percentile at the end-of-kindergarten by between 8% and 11%. But, if we instead provide children what the high income 90th percentile households provide in home investments, the 90-10 gap closes by substantially more, between 16% and 27%. Our finding of the importance of home environments to skill gaps by income echoes recent findings that long-term gaps in test scores by SES have failed to close in the United States, despite substantial increases and re-distribution of public school funding (Hanushek et al., 2019a,b).

Another important finding is that classroom quality has larger effects on the children entering kindergarten with low skills, essentially heterogeneous "treatment effects." The effect of classroom investments in children are 1.44 to 1.95 times higher for children in the lowest decile of initial skills than for high skill top decile children. This finding indicates that there are substantial gains in re-distributing classroom resources to target disadvantaged children.

Finally, we assess the importance of methodology to our conclusions, to provide

some guidance for future research on the importance of the generalities we allow. We find that, in general, not correcting for measurement error biases estimates substantially, although the sign of the bias varies across parameters given the general non-linear model we estimate. In general, we find that importance of classroom (relative to home) is upwardly biased in models without measurement error corrections and in models not allowing for complementarities/interactions.

We also use our estimated model to compare the distributions of classroom effects/quality implied by our model to alternative value-added models. We find that the estimated classroom effects in this class of models are biased because of relevant omitted information about parental investments and children's non-cognitive skills. Our results suggest that, within our sample, the value-added models systematically confound the treatment effect of classroom with the student selection into classrooms based on their non-cognitive skills and their home quality. Additionally, we study if the systematic bias in the estimated classroom quality would jeopardize the validity of the value-added model as a method for teacher evaluation. We find that, because of the bias in classroom effects, the value-added model would systematically predict a upward biased evaluation for older teachers and a downward biased evaluation for black teachers, which dissipates as those teachers gain experience. However, the magnitude of the bias is not large on average.

Subsequent sections are organized as follows. Section 2 discusses the econometric model, identification assumptions, and estimation strategy. Section 3 describes the data and presents some descriptive patterns in development over the kindergarten year. Sections 4-5 reports the estimates and validation exercises. Sections 6-7 use the estimated model to assess standard valued-added models and decompose the determinants of child development. Section 8 concludes.

2 Model and Estimation Framework

This section presents our general framework for child development. Our goal is a general enough model of skill formation that it can nest the key components of the previous research. We discuss various conditions for identification of the model, and close the section by developing a practical multi-step estimator for the model that we can take to the data.

2.1 Skill Development

At each age $t = 0, 1, \ldots, T$ children are characterized by a set of J skills. Let $\theta_{j,i,t}$ be child i's stock of skill j at age t. The collection of J skills for child i is represented

by the vector $\Theta_{i,t} = \{\theta_{1,i,t}, \ldots, \theta_{J,i,t}\}$. Skills include both cognitive and non-cognitive skills. Skill j in the next period is produced according to this technology:

$$\theta_{j,i,t+1} = f_{j,t}(\Theta_{i,t}, H_{j,i,t}, C_{j,i,t}, \eta_{j,i,t}) \tag{1}$$

where $H_{j,i,t}$ is a vector of investments from home and $C_{j,i,t}$ is a vector of investments from the classroom the child attends.[2] The investments and skill stocks are all strictly positive. The child development function in equation (1) is indexed by j and t to emphasize that the technology itself is heterogeneous across skills and over ages. The unobserved shock in skill formation is defined as $\eta_{j,i,t}$, and we return to its properties below.

Investment from home represents all child development activities outside of school. This need not be solely from interactions with parents, but could involve non-parental caregivers such as after school care. Investment from the classroom can be from any interaction during the school day, including from teachers, other schools staff, and peers. In our empirical specifications we allow for classroom specific investment, which can expose children within the same school to different classroom level environments.

In the child development literature (see perhaps Heckman and Masterov, 2007), equation (1) is typically labeled a "skill production technology." In the education literature (see for example Rivkin et al., 2005; Krueger, 1999), equation (1) is labeled as a "education production function." In the former case, the skills include cognitive and non-cognitive skills measured in survey data, and the investments from parents are the focus of the analysis. In the latter case, reading or mathematics skills are typically assumed to be directly measured via standardized tests administered in schools, and the productivity of school inputs (teachers and other classroom attributes) is the focus. Our specification nests both of these frameworks.

2.2 Measurement

Our skill development/education production function (1) is written in terms of latent variables. We recognize that children's skills and the various investments in a child's skills from parents and classrooms are unobserved and only imperfectly measured in data. For this reason, we consider a measurement system which can incorporate several previous approaches to this issue.

[2]This formulation of the skill technology specifies skills as a first order Markov process. This is consistent both with the education value-added literature and the models developed by (such as work by Cunha and Heckman, 2007; Cunha et al., 2010). The formulation rules out longer lagged persistence as in Todd and Wolpin (2003).

First, we allow for multiple measures of latent variables, and following Cunha and Heckman (2007); Cunha et al. (2010), we conceptualize each measure M as imperfect and including measurement error. More formally, let $\omega_{i,t}$ be a generic latent variable, e.g. $\omega_{i,t} = \theta_{j,i,t}$. For each latent variable, we have $m = 1, 2, \ldots, K_{\omega,t}$ measures. The number of measures can vary across latent variables and periods, and depends on the available data. Each scalar measure is denoted $M_{\omega,i,t,m}$ and takes the form:

$$M_{\omega,i,t,m} = b_{\omega,t,m}(\omega_{i,t}, \epsilon_{\omega,i,t,m}), \tag{2}$$

where $\epsilon_{\omega,i,t,m}$ is the measurement error for measure m and $b_{\omega,t,m}$ is the measurement function. One of the key identification issues is then how the measurement error ϵ relate to the latent variable ω. In much of the previous literature, it is simply assumed that there is no measurement error at all, and the particular measures at hand are exact measures of the associated latent variable: $M_{\omega,i,t,m} = \omega_{i,t}$. Our framework generalizes this approach.

A second measurement issue concerns the latent classroom investments. One could take a similar approach with classroom inputs. In particular, we could assume that observed classroom characteristics, such as class size or teacher experience, are imperfect measures of classroom quality (see for example Bernal et al., 2016). In that case, we could implement the same measurement system as in (2) to recover latent classroom investments. Instead, we take a different, and more general, approach, and treat the classroom effects in line with the education literature, as latent fixed effects. Exploiting the clustered survey design of our data, which surveyed multiple students per classrooms, we estimate the distribution of classroom quality via a generalized non-linear fixed effect estimator. We detail this approach below.

2.3 Baseline Empirical Specification

The model presented above provides some of the general concepts of our empirical specification. Next we present specific functional forms, identification assumptions, and our estimation strategy, which we can take directly to data. We start with a baseline specification, based on a particular specification of the production technology (1) and measurement system (2). This specification is the most restrictive specification we consider, and we generalize it in subsequent sections.

The baseline specification assumes a log-linear, Cobb-Douglas, form for the production technology. We specify the skill development function for each skill j (1) as

$$\ln \theta_{j,i,t+1} = \ln A_{j,t} + \gamma_{1,j} \ln \theta_{j,i,t} + \gamma_{2,j} \ln C_{j,i,t} + \eta_{j,i,t} \,. \tag{3}$$

Skill j in period $t+1$ is produced by the previous period stock of that skill $\theta_{j,i,t}$ and classroom investment in that skill $C_{j,i,t}$. To save on notation, we do not include a classroom or school subscript, but the model allows for clustering of classroom investments at the classroom and school level. The parameter $\gamma_{1,j}$ provides the relative productivity of the existing stock of skills. The productivity of classroom investments is captured by $\gamma_{2,j}$. The skill production shock is defined as $\eta_{j,i,t}$, which captures unobserved inputs in child development, and $\ln A_{j,t}$ represents total factor productivity (TFP). We normalize the shocks $\eta_{j,i,t}$ to be mean zero for all j, t, implying that the TFP term $\ln A_{j,t}$ can be thought as the mean of the "general" structural shock $(\ln A_{j,t} + \eta_{j,i,t})$.

As in much of the previous literature, we consider a linear (or log-linear) system of measures for the latent skills stocks. For each latent variable $\omega_{i,t} \in \{\theta_{1,i,t}, \ldots, \theta_{J,i,t}\}$ and period t, we have $m = 1, 2, \ldots, K_{\omega,t}$ measures given by

$$M_{\omega,i,t,m} \quad = \quad \mu_{\omega,t,m} + \lambda_{\omega,t,m} \ln \omega_{i,t} + \epsilon_{\omega,i,t,m}, \tag{4}$$

where $M_{\omega,i,t,m}$ is the mth specific measure, $\epsilon_{\omega,i,t,m}$ represents the measurement error, and $\mu_{\omega,t,m}$ and $\lambda_{\omega,t,m}$ are the measurement parameters. $\mu_{\omega,t,m}$ and $\lambda_{\omega,t,m}$ provide the location and scale of the measure m. The parameter $\lambda_{\omega,t,m}$ is the factor loading for the latent factor $\omega_{i,t}$ and measure m. Given the inclusion of the intercept, we normalize the measurement error to be mean-zero without loss of generality: $E(\epsilon_{\omega,t,m}) = 0$ for all ω, t, m. In our data, the set of measures for each type of children's skills $\omega_{i,t} \in \{\theta_{i,1,t}, \ldots, \theta_{i,J,t}\}$ is a combination of available assessments, as we discuss in more detail below.

2.4 Identification of Baseline Specification

Next, we describe the identification of our baseline specification in (3). The concepts introduced here also apply to the identification of the more general models that we explore next, but are more easily discussed in a simplified setting.

The baseline specification in (3) is identified up to some initial normalization given that latent skills (generically indexed by ω) and classroom inputs (C) are not directly observed and have no particular location or scale. We normalize all of the initial period ($t = 0$) latent variables to be mean 0 and variance 1:

Normalization 1 *Initial period ($t = 0$) normalizations:*
(i) $E(\ln \omega_{i,0}) = E(\ln C_{j,i,0}) = 0$
(ii) $V(\ln \omega_{i,0}) = V(\ln C_{j,i,0}) = 1$
for all $\omega_{i,0}$ *and* j.

With these normalizations, we treat all latent variables symmetrically, imposing the same normalizations on each in order to ease the interpretation of the estimates.[3] The normalization resolves the arbitrariness of the measures. Any positive monontonic transformation of the measure would also be a valid measure.

One normalization that appears non-standard relative to the prior literature is that we write the classroom effect as $\gamma_{2,j} \ln C_{j,i,0}$, because of the scale normalization in Normalization (1). That is, given the normalization that $V(\ln C_{j,i,0}) = 1$, the standard deviation of the classroom quality distribution is equal to the parameter $\gamma_{2,j}$. This normalization of the classroom effects implies that all of the technology parameters in (3), which represent the productivity of each factor in producing a child's skills, can easily be compared.

Next we move to the identification analysis. Given the normalizations on the location of the latent skills, the measurement intercepts are identified from the mean of the observed measures:

$$\mu_{\omega,0,m} = E(M_{\omega,i,0,m}) \text{ for all } \omega_{i,0} \in \{\Theta_{i,0}\}. \tag{5}$$

We cannot identify the scaling parameters for the initial period, $\lambda_{\omega,0,m}$, without further restrictions on the measurement errors. We consider the following independence assumptions, commonly used in this literature (e.g. Cunha et al., 2010), that the measurement errors are independent of each other and of the latent variables:

Assumption 1 *Measurement model assumptions:*
(i) $\epsilon_{\omega,i,t,m} \perp \epsilon_{\omega,i,t,m'}$ *for all* t, $m \neq m'$, *and latent variable* ω
(ii) $\epsilon_{\omega,i,t,m} \perp \epsilon_{\omega',i,t,m}$ *for all* t, m, *and latent variable* $\omega \neq \omega'$
(iii) $\epsilon_{\omega,i,t,m} \perp \epsilon_{\omega,i,t',m'}$ *for all* $t \neq t'$, *all* m *and* m', *and latent variable* ω
(iv) $\epsilon_{\omega,i,t,m} \perp \omega'$ *for all* t, m, *and latent variables* $\omega \neq \omega'$

Assumption 1 (i) is that measurement errors are independent contemporaneously across measures. Assumption 1 (ii) is that measurement errors are independent contemporaneously across different measures of different factors. Assumption 1 (iii)

[3]Note that we do not impose any restrictions on the latter period ($t > 0$) latent variables, for example the stock of latent skills in periods $t > 0$. See (Agostinelli and Wiswall, 2016b) for an analysis the potential biases caused when latent variables are normalized in all periods of dynamic model.

is that measurement errors are independent over time. Assumption 1 (iv) is that measurement errors are independent of the latent variables for child skills and investments. Although these assumptions are strong in some sense, they are common in the current literature.[4]

Under these assumptions, the initial period ($t = 0$) scaling factors are identified from ratios of covariances between the measures:

$$\lambda_{\omega,0,m} = \sqrt{\frac{Cov(M_{\omega,0,m}, M_{\omega,0,m'})Cov(M_{\omega,0,m}, M_{\omega,0,m''})}{Cov(M_{\omega,0,m'}, M_{\omega,0,m''})}} \tag{6}$$

for any three measures $m \neq m' \neq m''$.[5] Given the identification of the measurement parameters for the initial period, we identify latent variables up to the measurement errors:

$$\ln \omega_{i,0} = \frac{M_{\omega,i,0,m} - \mu_{\omega,0,m}}{\lambda_{\omega,0,m}} - \frac{\epsilon_{\omega,i,0,m}}{\lambda_{\omega,0,m}} = \widetilde{M}_{\omega,i,0,m} - \frac{\epsilon_{\omega,i,0,m}}{\lambda_{\omega,0,m}} \tag{7}$$

where $\widetilde{M}_{\omega,i,0,m}$ is the transformed measure using the identified measurement parameters.

To analyze the identification of the skill development function (3), we substitute the transformed measures for both periods $t = \{0, 1\}$ skills (7) into (3), and after some algebra, the empirical analogue of the technology of skill formation can be written as:

$$M_{j,i,1,m} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} + \lambda_{j,1,m}\gamma_{1,j}\widetilde{M}_{j,i,0,m} + \lambda_{j,1,m}\gamma_{2,j} \ln C_{j,i,0} + \kappa_{j,i,0,m}\,, \tag{8}$$

where the residual $\kappa_{j,i,0,m}$ is given by

$$\kappa_{j,i,0,m} = \lambda_{j,1,m}\eta_{j,i,0} - \gamma_{1,j}\frac{\lambda_{j,1,m}}{\lambda_{j,0,m}}\epsilon_{j,i,0,m} + \epsilon_{j,i,1,m}\,. \tag{9}$$

[4]Our assumption of full independence is sufficient, but not necessary, for at least some of our identification analysis. Below, we point out instances where weaker assumptions, allowing for some forms of dependence among measures and among measures and latent variable, can be used for identification. In cases of parametric production functions and parametric measurement system, we can relax that assumption with weaker conditions, for example assuming zero correlation.

[5]As we show in the Appendix, in the presence of at least two latent skills, the identification of the measurement systems is possible with only two measures for each latent skill. This relies on the independence of measurement errors across latent variables.

Note that, at this point, we have not identified the measurement parameters for period 1 ($\mu_{j,1,m}$ and $\lambda_{j,1,m}$), only those for period 0. We do not want to impose any restrictions on both periods because these would generally imply restrictions on the skill development process (Agostinelli and Wiswall, 2016b).

The equation in (8) can be re-written in "reduced form" as

$$M_{j,i,1,m} = \beta_{0,j} + \beta_{1,j}\widetilde{M}_{j,i,0,m} + \beta_{2,j} \ln C_{j,i,0} + \kappa_{j,i,0,m} \tag{10}$$

where the set of reduced form parameters $(\beta_{0,j}, \beta_{1,j}, \beta_{2,j})$ are functions of both structural (production function) and measurement parameters:

$$\beta_{0,j} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} \tag{11}$$

$$\beta_{1,j} = \lambda_{j,1,m}\gamma_{1,j} \tag{12}$$

$$\beta_{2,j} = \lambda_{j,1,m}\gamma_{2,j} \tag{13}$$

The system in (11)-(13) includes 3 equations for 5 unknowns. Agostinelli and Wiswall (2016a) show that this under-identification problem can be solved in the presence of age-invariant measures for skills. A pair of measures $M_{t,m}$ and $M_{t+1,m}$ for latent variables ω_t and ω_{t+1} is *age-invariant* if $E(M_{t,m}|\omega_t = p) = E(M_{t+1,m}|\omega_{t+1} = p)$ for some $p \in \mathbb{R}_{++}$. Intuitively, age-invariance implies that the expected measure M, say a test score, for two children of different ages but equal skills is the same. This rules out age specificity of measures between ages t to $t+1$. And, this assumptions implies that the measurement parameters are constant with respect to a child's age ($\mu_{j,1,m} = \mu_{j,0,m}$ and $\lambda_{j,1,m} = \lambda_{j,0,m}$). In the Data Section, we discuss the measures in our particular dataset and whether the age-invariance assumption is appropriate. To identify our model, we require that not all measures are age-invariant, but that we have at least one:

Assumption 2 *The data contains at least one age-invariant measure for each skill j.*

Assumption 2 allows us to identify the structural parameters from the reduced-form parameters as

$$\ln A_{j,0} = \frac{\beta_{0,j} - \mu_{j,0,m}}{\lambda_{j,0,m}} \tag{14}$$

$$\gamma_{\ell,j} = \frac{\beta_{\ell,j}}{\lambda_{j,0,m}} \quad \forall\, \ell \in \{1,2\} \tag{15}$$

where the measurement parameters for the initial period ($\mu_{j,0,m}$, $\lambda_{j,0,m}$) are already identified, up to a normalization, as shown in (5) and (A-2).

Two main identification challenges remain. Equations (14) and (15) show that in order to identify the structural parameters we need to consistently estimate the reduced-form parameters βs. The first issue is measurement error. In (10), the right-hand side measure of children's skills $\tilde{M}$ are correlated with the residual error κ because κ includes the measurement error ϵ associated with $\tilde{M}$. Even ignoring the unobservability of the classroom fixed effect, the OLS estimator of β_1 would be biased by measurement error, and in the simplest linear case with no other covariates, the OLS estimator of β_1 would be attenuated toward 0.[6]

The second identification challenge comes from selection bias with respect to the structural component η of the residual κ. If children systematically sort into classrooms based on unobserved characteristics/inputs η, we would confound the classroom effect on child development with these baseline characteristics. Similarly, a bias exists if children's skills at entry are correlated with the unobservable characteristics η. In this restricted model, η includes any home influences on skill development, and it is plausible that the home environment is related systematically to the chosen school and classroom environment and the child's kindergarten skills at entry.

The main identification assumption of the baseline model is based on the conditional (on latent child skill) exogeneity of the unobserved shock. We formalize the identification assumption as follows:

Assumption 3.a *Mean-independence of the production function shock:*

$$E(\eta_{j,0}|\theta_{j,0}, C_{j,0}) = 0$$

This assumption is similar to the typical one for fixed effect models, but, in our case, the classroom fixed effects and the $\theta_{j,0}$ variable are unobserved. As in the standard fixed effect case, this assumption allows for unrestricted sorting into classroom/schools based on unobserved child skills $\theta_{j,0}$ (say disadvantaged children attend low quality classrooms). In subsequent models, we weaken this assumption by generalizing the model and including more conditioning variables (additional latent skills, home investments, and interactions/complementarities).

Under Assumption 3.a, the reduced-form parameters βs are consistently identified using the multiple excluded measures as instrumental variables to adjust for measurement error (see Agostinelli and Wiswall, 2016a).[7]

[6]In this specification, with the inclusion of classroom fixed effects, we may not be able to sign the bias in this way. The measurement error bias becomes even more difficult to sign in the more general specifications we consider later, which include additional covariates and non-linearities.

[7]Note for this simple linear model, we can use standard within classroom differencing to elim-

2.5 Generalizing: Parental Investments

We generalize our baseline specification in (3) by including parental investments. This specification includes a unidimensional parental investment input which is common across the different skill domains, but is allowed to be differently productive in producing math, reading and non-cognitive skills. Specifically, the technology of skill formation for each type of skill j is

$$\ln \theta_{j,i,t+1} = \ln A_{j,t} + \gamma_{1,j} \ln \theta_{j,i,t} + \gamma_{2,j} \ln C_{j,i,t} + \gamma_{3,j} \ln H_{i,t} + \eta_{j,i,t} \tag{16}$$

where $H_{i,t}$ is the parental investments, and the parameter $\gamma_{3,j}$ is the productivity of parental investments in producing skill j. We assume that parental investments $H_{i,t}$ are unobserved and imperfectly measured. We allow for multiple proxies for investments in the data and for measurement error, in the same fashion as for children's skills:

$$M_{H,i,t,m} = \mu_{H,t,m} + \lambda_{H,t,m} \ln H_{i,t} + \epsilon_{H,i,t,m} \tag{17}$$

where we maintain the assumptions on the measurement errors as in Assumption (1). We generalize Assumption 3.a by conditioning on latent home investments:

Assumption 3.b *Mean-independence of the production function shock:*

$$E(\eta_{j,0} | \theta_{j,0}, H_0, C_{j,0}) = 0$$

Following the identification discussion in section (2.4), we can write the technology in (16) in terms of the measures as

$$\begin{aligned} M_{j,i,1,m} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} + \lambda_{j,1,m} \gamma_{1,j} \widetilde{M}_{j,i,0,m} + \\ \lambda_{j,1,m} \gamma_{2,j} \ln C_{j,i,0} + \lambda_{j,1,m} \gamma_{3,j} \widetilde{M}_{H,i,0,m} + \kappa_{j,i,0,m} \end{aligned} \tag{18}$$

where the structural parameters of interested can be identified under assumption 3.b and age-invariance as shown in equation (14)-(15).

inate the classroom effect. In subsequent models including interaction terms, estimation is more involved, as we detail below.

2.6 Generalizing: Multiple Skills and Complementarities

Next, we generalize the technology in (16) in two ways: (i) we include all the J skills in the technology of skill formation for each type of skill j; and (ii), we allow for complementarities between classroom investments and skills. In other words, we allow for heterogeneity in the productivity of class investments with respect to initial stock of skills. Following Agostinelli and Wiswall (2016a), we consider a kind of trans-log technology of skill formation with interaction terms between investments and skills. The technology for each type of skill j is

$$\ln\theta_{j,i,t+1} = \ln A_{j,t} + \sum_{k=1}^{J} \gamma_{1,j,k} \ln \theta_{k,i,t} + \gamma_{2,j} \ln C_{j,i,t} + \gamma_{3,j} \ln H_{i,t} + \gamma_{4,j} \ln \theta_{j,i,t} \ln C_{j,i,t} + \eta_{j,i,t} \tag{19}$$

where the parameters $\{\gamma_{1,j,k}\}_{k=1}^{J}$ represent the elasticity of skills $\theta_{j,i,t+1}$ with respect to each type k of skills ($\theta_{k,i,t}$). The TFP term $\ln A_{j,t}$ represents the intercept of the model, and in the empirical analysis will be also function of children's observable characteristics like gender and age. The parameter $\gamma_{4,j}$ governs the complementarity between classroom investments and skills. Here the elasticity of classroom investments depends on the child i's skills at entry:

$$\frac{\partial \ln \theta_{j,i,t+1}}{\partial \ln C_{j,i,t}} = \gamma_{2,j} + \gamma_{4,j} \ln \theta_{j,i,t} \tag{20}$$

$\gamma_{4,j} > 0$ implies a higher return to classroom investment for children with high initial skills relative to children with low initial skills. In contrast, $\gamma_{4,j} < 0$ implies a higher return of classroom investments for children with low initial skill. The sign of this parameter indicates how policy interventions (e.g. improved classroom quality) should be targeted to maximize skill development. We note that an alternative form of the production function, a Constant Elasticity of Substitution (CES) function,

$$\theta_{j,i,t+1} = A_{j,t}(\sum_{k=1}^{J} \gamma_{1,j,k}\theta_{k,i,t}^{\sigma_j} + \gamma_{2,j}C_{j,i,t}^{\sigma_j} + \gamma_{3,j}H_{i,t}^{\sigma_j})^{1/\sigma_j} \exp(\eta_{j,i,t}),$$

with standard parameter restrictions, implies that each input is a weakly positive complement: $\frac{\partial^2 \theta_{j,i,t+1}}{\partial C_{j,i,t} \partial \theta_{j,i,t}} \geq 0$. This forces, by functional form, that the productivity of investments targeted to higher skill children would be larger than if targeted to lower skill children.

Because of the generality of the model, the identification of this model requires a weaker version of Assumption 3.b because we allow for non-linear and heterogeneous treatment effects of classroom by children's skills:

Assumption 3.c *Mean-independence of the production function shock:*

$$E(\eta_{j,0}|\theta_{1,0},\ldots,\theta_{J,0},H_0,C_{j,0}) = 0$$

The empirical model for the technology in (19) can be constructed by substituting the transformed measures $\widetilde{M}$ for each of the corresponding latent factors, as shown above. This generalized specification is the one we use for the main empirical analysis. Adding the interaction terms in this more general specification complicates the estimation because unobserved classroom quality interacts with the unobserved skills of children. Next, we explain how we address this problem.

2.7 Estimation

We conclude this Section by developing an algorithm for estimating the generalized technology of skill formation (19). The general concept of the estimation algorithm applies to several other versions of the model we estimate. The main estimation challenge is that the data contains only imperfect measures of skills and home investments, and classroom quality has no measures at all and is inferred from observing multiple students per classroom. In addition, although we treat the classroom effect as a general fixed effect, the simple within classroom transformation cannot be used because the classroom effects enter the model non-linearly (interacted with latent variables). For this reason, we take a different approach, and develop an estimation algorithm that accounts for both measurement error and interactions between the fixed effects and latent variables. Our iterative algorthm follows Arcidiacono et al. (2012), but unlike their case we use a 2SLS IV estimator to account for measurement error.[8]

[8]Note that we implement a multi-step algorithm to reduce the computational burden of the estimation. Alternatively, we could estimate the model in a single step via a GMM estimator, where the selected moments would include both the within-classroom orthogonality conditions, as well as the average within-classroom residual growth in skills for each classroom. This method would become more and more computationally intense as the number of classrooms increase (we need to recover a fixed effect for each classroom). In our case, given we have over 1,000 classrooms in our sample, the direct method would require us to simultaneously estimate well over 3,000 parameters (over 1,000 classroom fixed effects for each of the 3 types of skill technologies).

After estimation of the initial conditions and measurement parameters, following the identification arguments above, the estimation algorithm for the production function parameters takes two recursive steps. In the first step, we recover the classroom latent quality distribution for a given set of technology parameters. In the second step, we estimate a new set of technology parameters given the updated classroom quality distribution from the first step. We repeat this two-step sequence until convergence to a fix point for the parameters. The algorithm is described more formally as follows:

Algorithm 1 *We start with an initial $n = 0$ guess for the parameters:*

$$(\ln A^0_{j,t}, \left\{\gamma^0_{1,j,k}\right\}_{k=1}^{J}, \gamma^0_{2,j}, \gamma^0_{3,j}, \gamma^0_{4,j}).$$

For each iteration $n \in \{0, 1, ...\}$, we compute the following steps in sequence:

- ***Step 1.*** *Given the current parameter guess n, estimate the classroom fixed effect for each classroom as the average within-classroom residual in skills at the end of kindergarten:*

$$\ln C^n_{j,i,0} = \frac{\sum_{i' \in c(i)} \left[M_{j,i',1,m} - \ln A^n_{j,0} - \sum_{k=1}^{J} \gamma^n_{1,j,k} \widetilde{M}_{k,i',0,m} - \gamma^n_{3,j} \widetilde{M}_{H,i',0,m} \right]}{\sum_{i' \in c(i)} \left[\gamma^n_{2,j} + \gamma^n_{4,j} \widetilde{M}_{j,i',0,m} \right]}$$

  where $c(i)$ is the set of sampled children in the classroom that child i attends.

- ***Step 2.*** *Given the distribution of classroom effects $\ln C^n_{j,i,0}$ from Step 1, we estimate the following empirical analogue of the technology in (19)*

$$\begin{aligned}\widetilde{M}_{j,i,1,m} = &\ln A^{n+1}_{j,0} + \sum_{k=1}^{J} \gamma^{n+1}_{1,j,k} \widetilde{M}_{j,i,0,m} + \gamma^{n+1}_{2,j} \ln C^n_{j,i,0} + \gamma^{n+1}_{3,j} \widetilde{M}_{H,i,0,m} + \\ &\gamma^{n+1}_{4,j} \widetilde{M}_{j,i,0,m} \ln C^n_{j,i,0} + \kappa_{j,i,0,m}\end{aligned}$$

  where $\kappa_{j,i,0,m}$ is the error term, which includes both the structural shock $\eta_{j,i,0}$ as well as all the measurement errors for the measures of skills and home investments. As discussed above, OLS estimation of the Step 2 equation, even with the classroom effects known, would produce an inconsistent estimate of the

remaining parameters. We estimate the parameters using 2SLS using the multiple excluded measures of skills and parental investments ($M_{j,i,0,m'}, M_{H,i,0,m'}$, for some $m' \neq m$) as instrumental variables. This produces the $n+1$ iteration of parameters $(\ln A_{j,0}^{n+1}, \{\gamma_{1,j,k}^{n+1}\}_{k=1}^{J}, \gamma_{2,j}^{n+1}, \gamma_{3,j}^{n+1}, \gamma_{4,j}^{n+1})$*, which can be used for Step 1.*

The iteration procedure stops when all of the parameters converge, e.g. $||\gamma^{n+1} - \gamma^{n})||_{\infty} \approx 0$*. Otherwise we return to Step 1 with an updated set of parameters.*

The algorithm is relatively simple, fast, and performs well in Monte Carlo simulations reported in the Appendix. All standard errors and statistical tests are computed using a clustered bootstrap over the classroom and student data. For each bootstrap repetition, we repeat all of the estimation steps (initial conditions, measurement parameters, and the iterative production function estimation). Our inference procedure therefore accounts for the clustering of the data and the multi-step estimation.

3 Data

Our data is from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K). The ECLS-K surveys a US nationally representative cohort of children who enter kindergarten in 1998-99. Importantly for our analysis, the ECLS-K employed a multi-stage, cluster sampling design. For each sample school, the ECLS-K sampled multiple classrooms (when there were multiple classrooms) and multiple students per classroom. The ECLS-K links each student to their classroom, to their sampled classmates within that classroom, and to the classroom teacher. Each student in the ECLS-K was interviewed twice during their kindergarten year, once at the beginning of the year in the Fall, and again toward the end of the academic year in the Spring. Building on existing psychometric work, the ECLS-K fielded their own extensive assessments of the children in various domains at each survey period. In addition to assessing the students, the ECLS-K also surveyed the classroom teacher and the parents. From the parent survey, the ECLS-K developed extensive measures of the child's home life and parental interactions with their children. We discuss each of these measures in more detail below.

3.1 Skill Measures

The child interview component of the ECLS-K included an individually administered assessment of the child consisting of cognitive (reading and mathematics) and non-cognitive domains.

3.1.1 Cognitive Skills

The ECLS-K cognitive assessment battery has three subject areas: language and literacy, mathematical thinking, and general knowledge. We use the language and literacy assessments as our reading measures, and the mathematical thinking assessments as our mathematics measures. For each subject area, it has a two-stage design, a first-stage routing test, and based on the routing test score, an appropriate second-stage form, consisting of a subset of the total questions. This common setup is designed so that for a finite test taking time, children do not waste time answering questions that are too easy or too difficult. For the common routing test given to all children, the ECLS-K provides a raw score, a count of the number of items a child answers correctly. For the second stage test, ECLS-K provides an Item Response Theory (IRT) score, an estimate of the number of items that the child would answer correctly if she were to take all of the questions on all forms.[9] From the first wave Fall testing and the second wave Spring testing, the assessments were designed to be comparable across periods allowing researchers to track learning across time. There is strong support therefore for the age-invariance assumption (Assumption 2).

3.1.2 Non-Cognitive Skills

We use the teacher ratings of each student's non-cognitive skills derived from three measures: approaches to learning, self-control, and interpersonal skills. Each of these measures is the average of multiple items/questions. Teachers' responses to each item were on a 4 point scale. The approaches to learning scale measures behaviors that relate to the child's interaction with the learning environment, and includes six items that rate the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization. The self-control scale includes four items that indicate the child's ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to pressure from peers. The five interpersonal skills items rate the child's skill in forming and maintaining friendships, getting along with people who are different, comforting or helping other children, expressing feelings, ideas, and opinions in positive ways, and showing sensitivity to the feelings of others.

[9] We can view the IRT score as some parameterized function of the child's answers to the second stage items, where the parameters of this function are estimated. Any estimation error in the construction of the IRT score would then be part of the measurement error of our model. The two tests for each subject–routing and IRT–do not share common questions. And, although the routing test determines where a child starts on the second stage assessment, the child can go up or down in question difficulty from this initial placement.

3.2 Home Investment Measures

The home investment measures derive from the parent interview component of the ECLS-K. One parent per child (typically the mother) was interviewed, and the data collected included information on family structure and demographics, parental education and household income, and various aspects of the child's home environment and parent and child interactions. Our home investment measures are a combination of characteristics of the mother (e.g. mother's education), resources available to the child (e.g. books and computers), and parental time with children (e.g. reading with children).

In the Appendix, we show descriptive statistics for 20 various home investment measures, and their relationship with parental income, both unadjusted (raw) and adjusted for the child's age (in months). In general, home investments are increasing in parental income. The relationship with income is particularly strong for the reading and computer availability measures.

3.3 Sample

We include all classrooms that have at least five children who have at least two non-missing mathematics, reading, and non-cognitive assessments at the beginning and at the end of the kindergarten year (12 total skill measures), and two measures of home investment during the kindergarten year. We drop students who switch schools or classroom during the kindergarten year because we cannot assign them a unique classroom. These criteria select 8,656 children within 1,118 classrooms and 637 schools. The Data Appendix shows the characteristics of our sample relative to the full ECLS-K sample. In general, our sample of children with complete information is more socio-economically advantaged relative to the full sample, as indicated by the higher household income in our sample of about $68,226 (2017 USD) in Table 1 compared to $62,432 USD in the full sample.

Table 1 presents descriptive statistics for our sample. At kindergarten entry, the average age of children is 5.68 years. Sixty-eight percent of the children are white, non-Hispanic, 14 percent are black, non-Hispanic, 9 percent are Hispanic, and 9 percent are of an other race/ethnicity. Sixty-nine percent of the children are living with both biological parents. Their mothers are on average 34 years old, have an average of 13.88 years of schooling, and work an average of 26.13 hours per week.

Examining the kindergarten classrooms and schools, the average class size is about 20 students. Some of the kindergarten classes are half-day, and others full-day. This implies that the average instructional time across all classrooms is 24 hours per week, lower than for later grades. The kindergarten teachers are overwhelmingly

female. They have an average of 9.53 years of experience in teaching kindergarten, and 35 percent of them have at least a master's degree.

Finally, about 31 percent of the schools in the sample are non-public schools, including secular and religious private schools. This distinguishes our sample (from a nationally representative survey) from other work relying on administrative records from public schools. We return to how these characteristics relate to classroom and school quality in later sections.

3.4 Child Development during Kindergarten

We briefly motivate our detailed econometric analysis by first summarizing patterns in child development during the kindergarten period. Figure 1 displays the average scores on the mathematics, reading, and non-cognitive assessments at the beginning and at the end of the kindergarten year, by the level of child's household income. In this Figure, each score is normalized using its respective mean and standard deviation at kindergarten entry to facilitate interpretation. The Figure reveals the wide dispersion of mathematics, reading, and non-cognitive scores at kindergarten entry across income deciles. At entry, the gap in average mathematics, reading, and non-cognitive scores between children at the lowest and the highest income decile is 1.1, 1.0, and 0.5 standard deviations, respectively.

Figure 2 plots the change in average scores between the beginning and the end of the kindergarten year. The figure indicates that, by the end of kindergarten, the dispersion of average mathematics and non-cognitive scores widens by 0.1 of a standard deviation, while the dispersion in reading score remains unchanged.[10] That the skill gaps by income do not fall through the kindergarten year, and for two domains actually increase, suggests that allocation of school and home investments is actually exacerbating the degree of inequality. Understanding the source of these patterns motivates our analysis.

4 Estimates

This section presents our estimates of the measurement model and production function components of the model. The initial conditions–the sorting of children to classrooms and distributions of home investments–we describe in a later section. Also in

[10]The gap in average mathematics, reading, and non-cognitive scores between 10th and 90th percentiles at the end of the kindergarten are 1.2, 1.0, and 0.6 of a standard deviation, respectively. Information on the measures, and figures for additional measures are available in the Appendix: the patterns are similar across the several other measures available.

later sections, we provide validation tests for the model, and interpret the estimates in a series of decomposition and counterfactual exercises.

4.1 Measurement Parameters

Table 3 presents the estimates of the measurement parameters for our measures of 4 latent variables: mathematics, reading, and non-cognitive skills, and home investments. (The fifth latent variable–classroom quality–has no particular measures.) The location and scale of each measure is arbitrary and is determined by the initial period normalization: each latent variable is normalized to be mean 0 and standard deviation 1. As described above, the location of a measure, given this initial normalization, is the mean value of the measure. For the mathematics routing test measure, a location of 5.11 indicates that, at the beginning of kindergarten, children score on average 5.11 on that test. Similarly, the measurement scale (sometimes referred to as the "factor loading") is identified from the correlation among the measures. Given the normalization that each latent variable has standard deviation 1, the scale can be interpreted as the effect of 1 standard deviation change in the latent variable. A value of 2.84 for the mathematics routing test indicates that an increase of a standard deviation in the latent factor predicts an average increase of 2.84 in this measure.

Table 3 also reports the signal-to-noise ratio for each measure at kindergarten entry. The signal-to-noise ratio $\in (0, 1)$ is the fraction of the variance of the measure that is explained by the latent variable. A higher signal-to-noise ratio indicates that the measure is more informative about the latent variable. The signal-to-noise ratio is identified from the correlation of the measures: a higher correlation across measures implies lower noise. Table 3 shows that the signal-to-noise ratios are high for our measures of mathematics and reading skills. For example, the signal-to-noise ratio of 0.93 for the mathematics routing test indicates that 93% of the measure's variation is due to the latent factor, while only 7% of the variance of the measure comes from noise.

The signal-to-noise ratio is lower for the non-cognitive measures, and in particular for the home investment measures. For the non-cognitive measures, this is perhaps not surprising. The non-cognitive measures are based on teacher observations, and perhaps more error prone than direct, test-based, student assessments of mathematics or reading. The home investment measures have signal-to-noise ratios ranging from 0.24 to 0.38. This too is perhaps not surprising given the difficulty of capturing all aspects of the home environment in a limited set of survey items. But even for these measures, the estimate of 25 percent signal provides some indication that these measures are certainly not entirely noise. We return to the issue of whether our

measures of the home environment are sufficient to capture relevant investments in a series of validation checks. In general, we find that our model is sufficiently rich enough to capture the influences of family income and other demographic variables without directly using these variables in the model.

That the measures have non-trivial error components is the main motivation for pursuing our measurement error correction procedure. We directly assess how measurement error biases estimates of the child development process below, by estimating models with and without the measurement error correction. For an initial indication of the importance of measurement error, note that in a simple linear regression of some outcome on a noisy measure with a signal-to-noise ratio of 0.5, the estimated slope coefficient would be biased toward zero, and only 1/2 of the true value. Although in our more complex non-linear models with multiple right hand side variables–all potentially measured with error–we cannot sign the bias, it is likely that measurement error is still an important consideration.

4.2 Production Technology

Table 4 presents the estimates of the baseline technologies, given in Equation (3) but augmented to include multiple skills. These are the simplest and most restricted models we estimate, and we provide estimates of several more general models next. The OLS panel of the Table shows the estimates not corrected for the measurement error, while the IV panel shows the estimates corrected for measurement error. Recall that our correction for measurement error is implemented using alternative measures as instruments for the main measure used. We estimate separate models for each of the three latent skills at kindergarten exit, mathematics, reading, and non-cognitive.

4.2.1 Prior Skills

Focusing on our preferred IV measurement error corrected estimates first (we discuss the importance of measurement error later), Table 4 indicates that all skills measured at kindergarten entry are important in the production of each of the skills at exit, but the magnitudes vary substantially. The estimate of 0.741 indicates that a 1 standard deviation increase in (log) mathematics skills at kindergarten entry produces (log) mathematics skills of 0.741 standard deviations. Note that the magnitude of this estimate depends on the initial normalization we chose: all initial period latent variables are normalized to be mean 0 and standard deviation 1. Given the log-log interpretation for the production technology, the estimate can also be interpreted as an elasticity: a 1 percent increase in latent mathematics skills at entry leads to

a 0.741 percent increase in mathematics skills at exit.[11] Reading and non-cognitive skills at entry have non-zero productivity in producing mathematics skills at exit, but the magnitude in each case is less than 0.1, indicating a strong specificity in skill. The remaining IV/measurement error corrected columns show a similar pattern, with stronger own-skill productivity than cross-skill productivity.

4.2.2 Classrooms

In addition to kindergarten skills at entry, these baseline models also include a classroom component, treated as a classroom specific fixed effect. As discussed above, we normalize the classroom effects to be mean 0 and standard deviation 1, as with all of the right-hand side latent variables. Therefore, there is a free parameter to be estimated on the classroom effect $\gamma_{2,j}$, and given the normalization, this parameter can be interpreted as the standard deviation of the classroom effects: $V(\gamma_{2,j} \ln C_{j,i,0}) = \gamma_{2,j}^2$.

Table 4 indicates that classrooms have a sizable effect on the production of the kindergarten skills. For example, the elasticity of end-of-kindergarten mathematics, reading, and non-cognitive skills with respect to classroom investments, after the measurement error correction, are 0.323, 0.381, and 0.519, respectively. These are sizable effects, as they are 40, 70, and 70 percent the corresponding effects for mathematics, reading, and non-cognitive skills at entry. This suggests that higher classroom investment can remediate a substantial fraction of the skill gaps at kindergarten entry. We return to this issue later, and directly compute a series of counterfactual re-sorting of students to classrooms in order to quantify the importance of classrooms.

4.2.3 Home Investment

We next turn to a second, and more general model, of skill development, including home/parental investments. We approach home investments in the same way as skills at kindergarten entry, and estimate models with and without measurement error corrections. As with all of the latent variables, home investments too are normalized to be mean 0 and standard deviation 1. Table 5 shows that home investments have a significant effect on child development, while the rest of the elasticities on prior skills remain unchanged. Focusing first on the measurement error corrected/IV estimates,

[11] In practice, in these translog functions, the log transformation of skills plays little role here except to make the model adhere to traditional production function concepts in which outputs and inputs are at least weakly positive. As in the previous literature employing factor models in this context, we assume the measures in the data, which can range over the whole real line, measure log inputs and outputs, rather than levels.

the elasticities of end-of-kindergarten mathematics, reading, and non-cognitive skills with respect to parental investments are 0.130, 0.094, and 0.083, respectively. These elasticities are approximately between 15% and 40% of the elasticities with respect to classroom investments, indicating the relatively higher productivity of the kindergarten classroom over the home investment. We emphasize that this is the difference in the relative productivity, and show below, in a series of counterfactual decompositions exercises, that the disparities in home environments are quantitatively more important than in those in school environments.

4.2.4 Complementarities and Heterogeneity

Table 6 shows the estimates for a third model, which includes a potential complementarity between the child's stock of skills and classroom investments.[12] Complementarity in these types of skill development models imply heterogeneity in the "treatment effect" of classroom investments: depending on the initial stock of skill entering kindergarten, children will experience different "returns" from a given classroom quality.

Heterogeneity by Initial Skill Table 6 documents a *negative* complementarity between initial skills and classroom investments: classroom investments are more productive for children with low initial skills relative to children with high initial skills. All of the interaction terms are statistically different from zero at standard levels. Our result here is in contrast to the typical positive complementarity assumption in which investments are more productive for higher skill children.

To better interpret the heterogeneity of the classroom effects in children, we graph the implied distribution of elasticities of classroom investments with respect to the stock of a child's skills in Figure 3. The figure demonstrates that the elasticities of classroom investments are decreasing in the child's skills at kindergarten entry. For example, a 1 standard deviation increase in classroom investments increases end-of-kindergarten mathematics, reading, and non-cognitive skills of a child in the lowest decile of math skills by 0.39, 0.46, and 0.59 standard deviations, respectively, whereas the same change in the classroom investments increases the skills of a child in the highest decile of initial math skills by 0.20, 0.25, and 0.41 standard deviations. The effect of classroom investments in children are 1.95, 1.84, and 1.44 times higher

[12]Estimates for even more general models are available in the Appendix. These include models where we allow for interaction between "cross-skills" and classroom, and skills and the home investment. The results are similar, and in particular, we estimate that the productivity of home investments is higher for lower skill endowed children.

for children in the lowest decile of initial skills than for high skill children. The heterogeneity of classroom effects with respect to reading and non-cognitive skills display a similar pattern.

Heterogeneity by Family Income Figure 3 also shows the heterogeneity of classroom effects with respect to household income. Household income is not directly part of the model, but it is positively correlated with the initial stock of a child's skills (as described more fully below). In this Figure, we compute the heterogeneous treatment effects conditional on income (Y), where average skill j conditional on income is given by

$$E(\theta_{j,i,1}|Y = y) = \int_{\Theta_0} E(\theta_{j,i,1}|\Theta_0) dF(\Theta_0|Y = y),$$

where $F(\Theta_0|Y = y)$ is the conditional distribution of the vector of latent initial skills with respect to income.

We estimate that a 1 standard deviation increase in classroom investments increases average end-of-kindergarten mathematics, reading, and non-cognitive skills of a child in the lowest decile of household income by 0.35, 0.43, and 0.55 percent, respectively, whereas the same increase in the classroom investments increases skills of a child in the highest decile of household income by 0.3, 0.35, and 0.50 percent, respectively. This implies that the elasticities of classroom investments for children in low-income households are 1.16, 1.22, 1.1 times larger than the elasticities for children in high-income households. These estimates suggest that targeting low income students with improved classroom quality would produce a higher effect on end-of-grade kindergarten skill development than similar investments given to students from high income households.

4.2.5 Importance of Measurement Error

The other key take-away from Table 4 is the importance of measurement error. Comparison between the OLS and the IV estimates shows that correcting for the measurement error increases own productivity (diagonal elements), but it decreases cross-elasticities (off-diagonal elements). For example, the elasticity of end-of-kindergarten mathematics skills with respect to initial mathematics skills increases from 0.599 to 0.741, while the elasticities with respect to initial reading and non-cognitive skills decrease from 0.133 and 0.139 to 0.064 and 0.094, respectively. These are sizable changes in the magnitudes of the productivity estimates, even with relatively high signal-to-noise measures, and suggest the importance of correcting for measurement

error. We return to the importance of measurement error again, when we describe how correcting for measurement error affects the counterfactual decomposition results we present below.

5 Testing and Validating the Model

In this section, we test our estimated model using various validation exercises. The intuition of these validation tests is that if our model is sufficiently rich enough to capture the key sources of child development, then the model should also be able to make valid out-of-sample type predictions. In particular, we show that our model can match patterns of child development by family income and race, even though our model does not include these variables. Our model is able to fully explain income and race gaps in child development using information about skills measured at kindergarten entry, parental investments, and kindergarten classroom quality. We also show that our estimated model can predict similar effects as the Tennessee STAR experiment. These results provide a kind of validation of the estimated model, and give us some confidence that the inferences derived from the estimated model do not suffer from serious omitted variable bias.

5.1 Validation I: Selection on Unobservables

The main assumption for our estimation exercise is essentially a model "sufficiency" one: our model, including children's skills entering kindergarten in 3 domains, home investments, and latent classroom quality (fixed effects), is rich enough to capture the key determinants of child development, so that the remaining unobservable dimensions are "ignorable." Formally, the "sufficiency" assumption is discussed in the Model and Estimation section. The threats to validity are the standard ones: that the included latent variables for initial skills, home, and school investments are correlated with any omitted aspects of child development.

We follow Chetty et al. (2014a) and use the rich set of "omitted" variables in our dataset, which are likely strongly correlated with any unobserved dimensions of child development, as the basis of a validation test. Strictly, the mean-independence assumption is untestable. But, as in Chetty et al. (2014a), we use household income as a proxy for these unobservable dimensions. As demonstrated in Figures 1-2, household income is strongly correlated with child skills at the end of kindergarten. However, household income is not included in our model, and therefore the variation from household income is a kind of out-of-sample variation we can use for testing.

We perform this test by regressing household income on the estimated residuals from our preferred, and most general, production function specification (estimates reported in Table 6). Table 7 shows that household income is unrelated with the residual variation of child development from our model, in each of the three domains (math, reading, and non-cognitive). We estimate several specifications, including no controls, with child demographic variables, and including the number of hours of mother's and father's labor supply. Across these specifications, we see that an increase of $100,000 in annual household income (an enormous change) is associated with effects from -0.02 to 0.02 standard deviations of skills, depending on the specification. These estimates are very small in magnitude, and are never statistically different from zero at conventional levels. As in Chetty et al. (2014a), we interpret this result as a failure to reject the null hypothesis of no omitted variable bias.

5.2 Validation II: Out-of-Sample Predictions

In a second test, we assess whether our estimated model is able to predict the observed variation in end-of-kindergarten scores by household income and race, variables which we omitted from the model. This idea is essentially the same as testing whether the estimated model can predict "untargeted" moments–an exercise commonly performed in structural or model based estimation. In this case, we are testing whether our model's specification based only on heterogeneity in children's initial skills, parental investments, and classroom quality is able to explain achievements gaps among socioeconomic groups.

Figure 4 shows the results for household income. The close match between our model prediction and the data indicates that the income gradient of children's skills is well explained by our model. And, Figure 5 shows similar results by the child's race–another variable excluded from our model.[13] This close match provides confidence that our model is not omitting important determinants of the child development process.

5.3 Validation III: the Tennessee STAR Experiment

Our final validation exercise uses an entirely different dataset and a randomized control treatment experiment, the Tennessee Student/Teacher Achievement Ratio

[13]This exercise would be uninformative if classrooms in our estimating sample were completely segregated by race. In that case, the model prediction would fit exactly the race patterns by construction via the classroom effects. As we show in Figure C-8, classes in our estimating sample are not fully segregated, but instead have some non-trivial variation in classroom racial composition.

(STAR) experiment. The STAR experiment was a four-year longitudinal study of the effects of class size in which over 11,000 students from 79 schools were randomly assigned into classrooms with different class size and teaching personnel. The experiment was initiated as the students entered kindergarten in the 1985-1986 school year and continued through third grade. [14] In one of the more recent evaluations of the results, Chetty et al. (2011), linking students to recent tax records, find that the students randomly assigned to higher quality kindergarten classrooms had higher adult earnings and college attendance rates. Their estimates indicate that improving class quality by 1 standard deviation increases annual earnings by $1,520 (9.6%) at age 27.

We use the STAR experiment to test the validity of our model. The intuition of this validation test is that if our estimated model provides a good representation of the child development process, with minimal omitted variable bias, then it should be also be able to replicate the STAR experimental results.

Matching the STAR Treatment Effect We first quantify, using our estimated model, what change in latent classroom quality would match the STAR experimental findings. This exercise is important for two reasons: (i) it gives us a first validation of the model; and (ii) it allows us to use our estimated model to extrapolate the STAR treatment effects for different changes in classroom quality. We replicate the STAR treatment effects at the end of kindergarten using the original STAR data.[15] And, we use our most general specification of our model (19), estimated using the ECLS-K data.

Figure 6 shows the treatment effects for math and reading test scores in the STAR experiment (dashed line) and how the different changes in latent classroom quality translate into changes in math and reading skills in our model (solid line). The solid line is a counterfacual manipulation of the classroom quality assigned to each student in our ECLS-K data, $\ln C + \delta$. Each point on the horizontal axis measures an increase in latent classroom quality provided to each student (δ), with $\delta = 0$ indicating the

[14] 6,323 children participated in STAR at kindergarten entry, and over 5,000 more started participation in later grades. STAR assigned children to one of three interventions: small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aide class (22 to 25 students with a full-time teacher's aide). See, for example, Krueger (1999) and Chetty et al. (2011) for a comprehensive summary of the experiment.

[15] Specifically, for each outcome Y, we estimate a model of $Y_{is} = \beta_0 + \beta_1 \text{class size}_{is} + \beta_2 \text{aid}_{is} + \alpha_s + \epsilon_{is}$, where class size_{is} is kindergarten class size for student i in school s, aid_{is} is a dummy variable for whether the class has a teacher's aid experimentally assigned, and α_s is a school specific fixed effect. We then instrument for class size using the randomly assigned treatment, 1 if small class or 0 if assigned a large class or large class with teacher's aid.

baseline/original level. The slope of this line comes from average estimated effect of classrooms on skill development, incorporating the estimated heterogeneous effects of classroom on students with different skill levels.

Where the STAR experimental estimate and our latent classroom simulation lines cross indicates the increase in latent classroom quality in our model that would exactly replicate the estimated STAR effect. In the case of math skills, the STAR treatment effect of a 5 percentile increase in end-of-kindergarten math skills is equivalent to assigning students in our estimated model $\delta = 0.55$ standard deviations higher latent classroom quality. For reading skills, the treatment effect of 6 percentiles is equivalent to $\delta = 0.5$ standard deviations higher latent classroom quality. Moreover, the solid line shows the changes in children's math and reading skills once we extrapolate the experimental results from STAR to different changes in classroom quality. Figure 6 indicates that an increase in $\delta = 1$ standard deviation in classroom quality would produce an increase in 9 percentiles for math skills, and 12 percentiles for reading skills.

Matching the STAR Latent Classroom Distribution As discussed by Chetty et al. (2011), the STAR experiment not only provides the effect of the particular class size experiment, but also exogenous variation in classroom assignment (within schools, where the randomization of classrooms took place), providing credible estimates of the effect of an exogenous change in general classroom quality on test scores. Next, we use this aspect of the STAR experiment and compare the "effect size" of classroom quality in the STAR experiment to the analog in our model. Here we are testing whether our model estimates, derived from including controls for latent child skills and home investments, can replicate the STAR experiment estimates.

Panel A in Table 8 shows the results for this validation exercise. We perform the validation exercise for both math and reading standardized test scores.[16] Columns (3) and (4) show the results for a simple classroom fixed effects model using STAR data, and Columns (1) and (2) show the estimated effects of latent classroom quality for model (19) with ECLS-K data. The two estimates are similar: the STAR and our estimates differ by 0.038 and 0.045 standard deviations. We take these results as suggesting limited bias in our estimates.

[16]Following common practice, we standardized (transform to mean 0, standard deviation 1) the end-of-grade scores in both datasets. Although the end-of-kindergarten measures in the STAR and ECLS-K data are different (an issue we return to below), we attempt to match statistics based on these standardized scores. It should be noted that standardizing test scores does not resolve the indeterminacy of the measurement model presented above, and our comparison here is only approximate.

Matching the STAR Heterogeneous Effects Our final validation exercises using the STAR experiment is to replicate the heterogeneity in treatment effects. A key finding from our analysis using the ECLS-K data is that the effect of a higher quality classroom is larger for those children entering kindergarten at a disadvantage in math, reading, and non-cognitive skills. Although the STAR dataset does not contain entry measures of skills, the estimated experimental treatment effects are generally larger for disadvantaged groups: low income students (those eligible for free or reduced price lunch) and black students.

Panel B in Table 8 compares the relative effects of the class size treatment on sub-groups defined by free/reduced price lunch and black and white race, for the STAR experiment and the analogous treatment using our estimated model.[17] In Panel B of Table 8, we compute ratios of treatment effects between two sub-groups. As derived in the Appendix, ratios of treatment effects are scale free measures, providing a statistic that is exactly comparable across the STAR and ECLS-K data, even though the two datasets use different end-of-grade measures. The cost of this procedure is of course statistical precision, as comparisons across sub-groups requires using smaller sub-samples.

Panel B in Table 8 shows that the relative effects of the class size treatment in both STAR and as predicted by our model are generally higher for the disadvantaged groups. In STAR, the class size effect on math and reading test scores for black children relative to white children are 1.08 and 1.20, respectively, implying that the effect of reducing class size is 8 and 20 percent higher for black children. Similarly, we find that decreasing class size has a 13 percent higher effect on reading skills for children who qualify for free or reduced-priced lunch relative to children who do not. On the other hand, for the same two subpopulations, we find that the relative treatment effect on math is slightly below one, although statistically indistinguishable from one.

Our estimated model using ECLS-K data replicates these relative classroom size effects fairly well. The model exactly replicates the relative classroom effects on math by race and on reading by free lunch. Moreover, the model predicts a similar higher relative black/white treatment effect for reading (both above 1). Finally, the model predicts a relative higher treatment effect of the children who qualify for free lunch on math, but this result differs from the point estimate in STAR (although

[17] Within our model, a change in class size is a change in classroom quality. Importantly, when comparing a common treatment effects between two groups, the relative effect does not depend on the magnitude of that treatment (no second order effects). Nevertheless, the non-linearities in the technology of skill formation can still generate heterogeneous treatment effects across subgroups if these subgroups differ in their initial skills.

the results are statistically the same). It should be noted that the standard errors (computed using a clustered by school bootstrap procedure) are large for the STAR estimates; unsurprising given the small sample sizes for the sub-groups. We perform a formal multiple hypothesis testing for math and reading separately, and fail to reject the hypothesis that the STAR treatment effects are equal to the ECLS-K values, with p-values of 0.84 and 0.95, respectively. With the precision caveat in mind, the fact that our model predicts similar heterogeneous effects as the randomized STAR experiment provides some additional confidence that our model estimates do not suffer from substantial omitted variable bias.

6 Assessing Value-Added Models

Previous education research makes use of administrative data from specific school systems to estimate classroom or teacher value-added. Although this data offers tremendous advantages due to the comprehensive coverage of all students in the school system, it typically only includes standardized mathematics and reading tests and limited information about other factors that we believe are important correlates with child development: the child's home environment and children's non-cognitive skills. In this section we quantify the biases in classroom value-added estimates that ignore home investments and non-cognitive skills.

For this analysis, we use our estimated model to compute the rank of each classroom in the estimated classroom value-added distribution derived from our nationally representative ECLS-K sample of kindergarten classrooms. Call this estimated classroom rank from the unrestricted model $\widehat{C}^{UR}_{Rank}$. We then compute a second estimate of the classroom value-added using our data, but a different, restricted, model, omitting the home investment and non-cognitive components, attempting to replicate a standard value-added type model.[18] Call the estimated rank computed using the standard VA model ($\widehat{C}^{VA}_{Rank}$).

Figure 7 shows the relationship between the standard VA model estimate ($\widehat{C}^{VA}_{Rank}$) and our unrestricted baseline estimates ($\widehat{C}^{UR}_{Rank}$). Although the rank estimates are strongly positively correlated, the Figure indicates that for both math and reading, there is considerable dispersion in the standard VA estimates, conditional on a given "true" estimate from our general model. For the classroom at the median of the unrestricted distribution, the restricted value-added models would produce classroom rank estimates ranging between the 10th and the 90th percentile and from the 20th and the 75th percentile, for math and reading skills, respectively.

[18] See Table C-5 for the full set of estimates for the VA model.

We directly quantify the difference in the two estimates by computing the following "bias" term:

$$Bias^{VA} = \widehat{C}^{VA}_{Rank} - \widehat{C}^{UR}_{Rank} \ ,$$

where $Bias^{VA}_{C} > 0$ (< 0) indicates that the standard VA model is overestimating (underestimating) the classroom value-added relative to our unrestricted model estimates.[19]

Figure 8 shows how the bias varies by the classroom's averagem parental investments and children's non-cognitive skills. As expected, larger (in absolute value) bias is associated with classrooms with higher average parental investments and higher average children's non-cognitive skills. In other words, we estimate that the omitted parental investments and children's non-cognitive skills are incorrectly "loaded" into the classroom's value-added in the standard VA models. The VA models would systematically predict a lower rank of classroom value-added for classrooms with lower level of parental investments and children's non-cognitive skills.

We also analyze how the bias in the value-added estimates relates to the characteristics of the classroom teacher. This analysis is important given a major goal of the value-added literature is the evaluation of teachers' productivity using these value-added estimates. We show that the estimates from the VA model can lead to systematic bias toward specific groups of teachers. Table 9 shows the regression results for the relationship between teacher's characteristics and the VA bias. The results in Table 9 indicate that a teacher's age and race are predictive of the VA bias, even within school, with an estimated upward bias for older teachers and a downward bias for black teachers, which dissipates as those teachers gain experience.

7 School and Family in Child Development

In this section, we use our estimated model to quantify the contribution of classroom and home investments in closing the achievement gaps between children from low and high income families. We begin our analysis by describing the estimated initial (ex ante or at kindergarten entry) joint distribution of latent classroom quality, home investments, and skills–the initial "sorting" of children to classrooms and home environments. We then specifically examine sorting by family income, which is not directly part of the model, but can be strongly related to the distribution of initial conditions. We conclude with an exercise in which we counterfactually manipulate

[19]Figure C-9 shows the marginal distribution of the value-added bias.

the initial conditions and quantify how these reallocations affect the level of ex post (at kindergarten exit) inequality.

7.1 Classroom, Home, and Initial Skills at Kindergarten Entry

We first use our estimated model to describe the "initial conditions" of the skill development model: the joint distribution of latent classroom quality, home investments, and latent kindergarten skills (at entry). It is important to note that these patterns cannot be directly examined using just descriptive correlations of data variables. The latent distributions we compute are "corrected" for measurement error using all of the available measurement data, and, in the case of the latent classroom quality, which has no direct measure, are inferred from the estimated outcomes.

Table 10 presents the estimates of the variance-covariance matrix of initial (at kindergarten entry) latent variables. Because the standard deviation of all inputs are normalized to one, covariances are equal to correlations. The estimates are derived from our most general model including latent skill and investment interactions, reported in Table 6.[20]

Table 10 shows that all three skill domains are positively correlated at kindergarten entry. Children who enter kindergarten with high mathematics skills also tend to have high skills in reading and in non-cognitive domains. Similarly, home investments received by the children during the school year are positively correlated with their initial skills, suggesting that there are "permanent" aspects of the home environment that give rise to highly developed children at entry and persist through kindergarten, reinforcing initial skill advantages.

Turning to the estimated sorting of students to classrooms, characterized by the latent classroom effect or value-added, we see a much weaker positive, and in some cases a weak negative, correlation between initial skills and classroom quality. Recall that we estimate classroom quality in each skill domain separately, allowing us to distinguish among classroom quality in mathematics, reading, and non-cognitive domains. Table 10 indicates that children with high mathematics and reading skills at entry have a slightly higher chance of receiving better quality classrooms in those domains. But that higher skill children at entry, across all domains, have a slightly lower chance of receiving a higher non-cognitive skill classroom.

[20]Note that in our framework, we impose no particular parametric assumption on this distribution, in contrast to some previous research which imposes that the marginal distribution of latent variables is Normally distributed or some low dimensional mixture of Normal distributions. Our estimate of the full joint distribution, beyond the correlations we report here, is available on request.

We can also estimate the relationship between classroom quality and latent home investment. Table 10 indicates that home and classroom quality are negatively correlated: children who attend a low quality classroom tend to have higher home investments. This provides some suggestion that home investments are compensating for lower quality classroom investments.[21] Greaves et al. (2019), for example, find similar results in the context of England. The authors find that improving (perceived) school quality causes parents to decrease the time investment into their children.

7.2 Inequality in Child Skills and Investments by Family Income

One of the early notions of public education in the United States and elsewhere is that public education would be an equalizing influence, exposing children of all income levels to a similarly high quality of education. However, the potential equalizing effect of public schooling is undone with tuition based private schools and locally financed public schools whose values are priced into housing, allowing higher income families to purchase higher quality schooling for their children.

We do not have an assumed model of how classroom sorting arises, but our estimates reflect, at least in part, how parents, constrained by available resources, choose neighborhoods and schools, and how children are assigned different classroom within schools.

Table 11 shows the relationship between our estimated latent classroom quality in mathematics, reading, and non-cognitive skills and the level of the child's household income. The Table demonstrates that classroom investments in mathematics and reading skills display a positive income gradient only for households with incomes below the sample median. That is, for the poorer households, higher income is correlated with higher quality classrooms, but for households with higher than median income, classroom quality is no longer related to income. We do not find any relationship between non-cognitive classroom investments and household income. In contrast, home investments are positively correlated with household income both for households below and above the sample median, but the relationship is stronger for

[21] This negative association between home investments and classroom quality is not driven by a mechanical relationship : because some kindergarten classrooms are all-day, rather than half-day, and these classrooms would tend to therefore have higher "investment," classroom investments may therefore be directly crowding out home investment. We repeat the results in Table 10 by controlling for the type of kindergarten: morning classes, afternoon classes and all-day classes. The results are unchanged. The results of this additional specification are available upon request.

the lower income households.

7.3 Decomposition Analysis

We next directly quantify the importance of the various elements of the model through a counterfactual decomposition analysis.

Unrestricted Model Panel A in Table 12 shows the effects for our baseline analysis. The first two columns show the average math, reading and non-cognitive latent skills at the end of kindergarten for the 10th and 90th percentile of the income distribution, respectively. The third column shows the achievement gap between the two subgroups (the difference between Columns 2 and 1). The 90-10 gaps are between about 0.7 and nearly 1, with respect to the initial latent distributions, normalized to have standard deviation 1. These gaps are in terms of latent skills, and as shown previously, our model estimates closely match the data distribution of observed measures.

In Columns (4) and (5), we counterfactually simulate changes in the initial conditions of the model. Column (4) shows the counterfactual achievement gaps predicted if all children were provided the average classroom quality of the high-income children, and everything else remains the same. In parenthesis we report the percentage change of achievement gaps relative to baseline (Column 3). Column (5) shows the counterfactual achievement gaps predicted from the model if the low-income children were provided with the average parental investments of the high-income children, and everything else remains the same.

Our results suggest that both classroom and home environments play an important role in explaining differences in a child development. Classroom quality closes between 4 and 9 percent of the skills gap, while parental investments close between 14 and 24 percent of the gap. This suggests that parental investments play a relatively more important role in explaining skills gaps between children from different socio-economic backgrounds. It is important to contrast this finding with the findings discussed above that latent classroom quality has a smaller effect-size on child skill formation than home investments. Home investments explain more of the income gap because home investments are more highly correlated with household income (i.e. more unequally distributed) than classroom quality, so that a counterfactual change in home investments is a larger investment change.

No Measurement Error Correction The previous results in Panel A were for the our full, preferred model. Panel B in Table 12 shows the effects for a restricted

version of the model, a restricted linear model where we do not correct for measurement error (i.e. estimated using OLS). Columns (4) and (5) show the decomposition results for this model. In this case, we find the opposite result as in Panel A: classroom quality is relatively more effective in reducing skills inequality.

The difference in the results between Panel A and Panel B suggests the importance of our empirical modelling choice. In particular, it highlights the importance of: (i) addressing the measurement error problem; and (ii) modeling heterogeneous classroom effects via our general interactions model. Only with the more general model do we see the importance of home environment as a key mechanisms in explaining developmental differences during kindergarten.

Value-Added Model Finally, Panel C estimates a third model, a simple VA specification, approximating a common specification in the prior education production function literature. In this specification, we omit non-cognitive skills, complementarities, and home investments, and do not correct for measurement error (i.e. estimated using OLS). As in Panel B, in this simple VA model we find that the classroom component is substantially more important than in our preferred estimates in Panel A, indicating that the simple VA model is biasing upward the importance of the classroom.

8 Conclusion

We develop an empirical framework that is general enough to nest many of the key features of two previously separate and parallel research programs, the Child Development literature and the Education Production Function literature. Our framework allows for both classroom and parental influences in child development, imperfect measures of both skills and inputs, cognitive and non-cognitive skills, and complementarities between children's skills and investments from home and school. We find that investments from classrooms and from home are important determinants of children's skills at the end of kindergarten. In addition, we document a negative complementarity between children's skills at kindergarten entry and investments from classrooms, implying that low-skill children benefit the most from an increase in the quality of schools. The counterfactual policy experiments show that providing all children with the 90th percentile of either classroom investments or home investments would substantially reduce the 90-10 skill gap, with greater effects for home investments because high quality home environments are more unequally distributed than school environments.

References

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. *NBER Working Paper*, (22442), 2016a.

Francesco Agostinelli and Matthew Wiswall. Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. *NBER Working Paper*, (22441), 2016b.

Peter Arcidiacono, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, 3:421–470, 2012.

Orazio Attanasio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. Estimating the production function for human capital: Results from a randomized control trial in colombia. *NBER Working Paper*, (20965), 2019a.

Orazio Attanasio, Costas Meghir, and Emily Nix. Human capital development and parental investment in india. *NBER Working Paper*, (21740), 2019b.

Pedro Bernal, Nikolas Mittag, and Javaeria A. Qureshi. Estimating effects of school quality using multiple proxies. *Labour Economics*, 39:1 – 10, 2016.

Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How does your kindergarten classroom affect your earnings? evidence from project star *. *The Quarterly Journal of Economics*, 126(4):1593–1660, 2011.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2014a.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–79, 2014b.

Flavio Cunha and James Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Ellen Greaves, Iftikhar Hussain, Birgitta Rabe, and Imran Rasul. Parental responses to information about school quality: Evidence from linked survey and administrative data. *Working Paper*, 2019.

Eric A. Hanushek, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. The achievement gap fails to close: Half century of testing shows persistent divide between haves and have-nots. *Education Next*, 2019a.

Eric A. Hanushek, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. The unwavering ses achievement gap: Trends in u.s. student performance. *NBER Working Paper*, (25648), 2019b.

James J. Heckman and Dimitriy V. Masterov. The productivity argument for investing in young children. *Review of Agricultural Economics*, 29(3):446–493, 2007.

James J. Heckman and Stefano Mosso. The economics of human development and social mobility. *NBER Working Paper*, (19925), 2014.

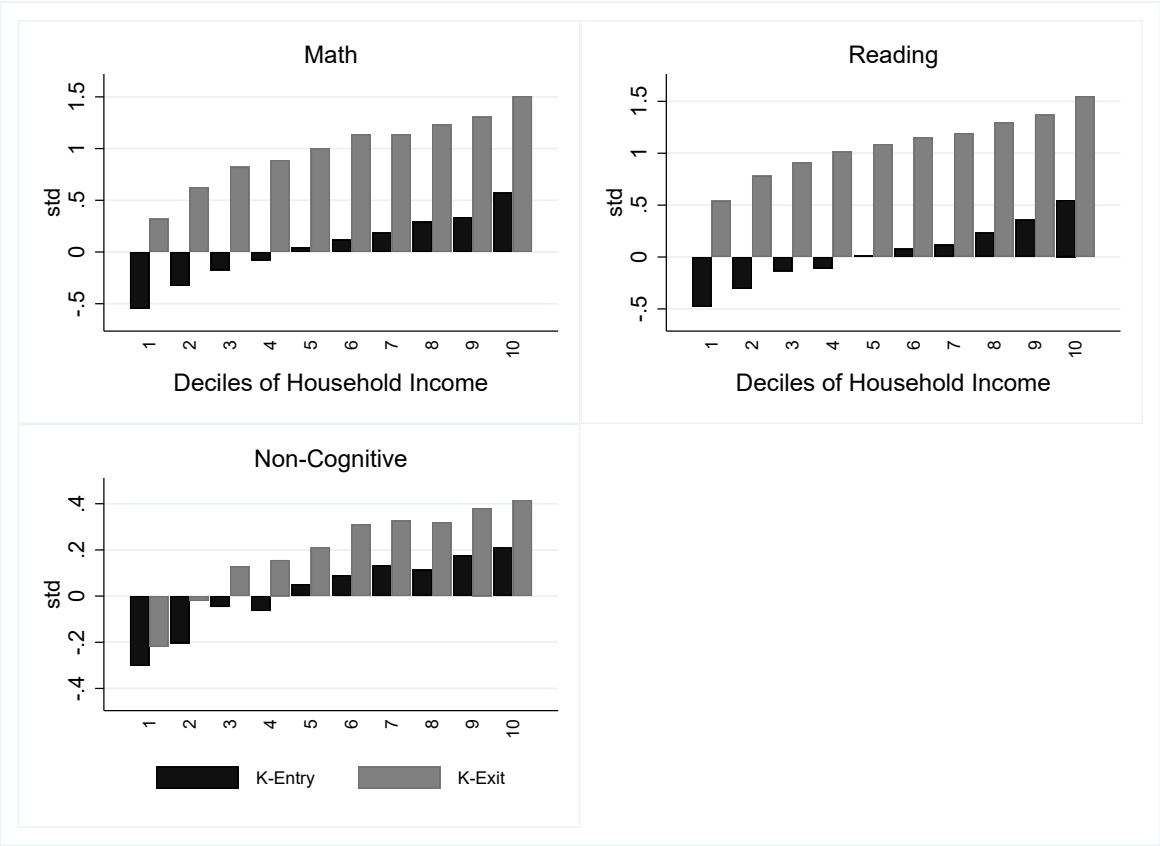
Kirabo C. Jackson. Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. *NBER Working Paper*, (18624), 2012.

Alan B. Krueger. Experimental estimates of education production functions*. *The Quarterly Journal of Economics*, 114(2):497–532, 1999.

Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Petra E. Todd and Kenneth I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485):F3–F33, 2003.

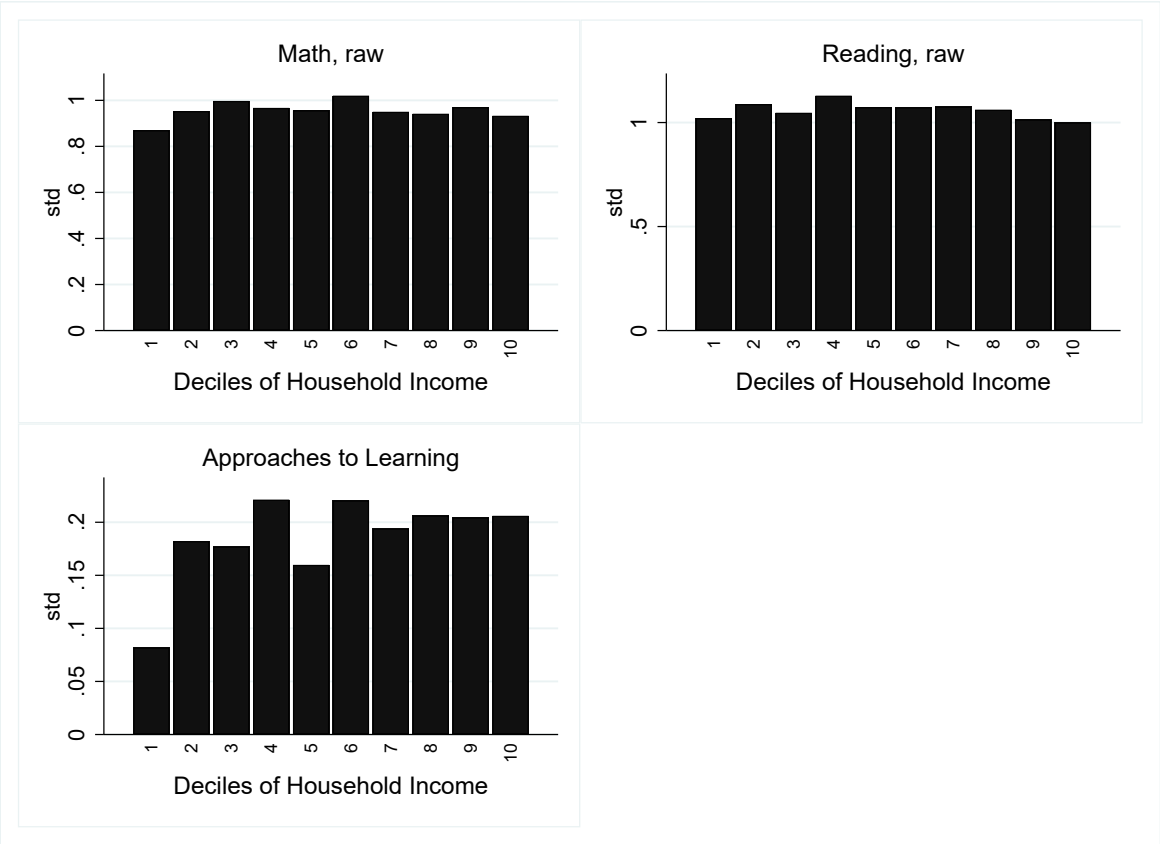
Figure 1: Cognitive and Non-cognitive Scores by Income Deciles



Source: ECLS-K:1998-99.

Notes: The Math and Reading scores are raw scores. The non-cognitive score is the approaches to learning score evaluated by the teacher. The scores are standardized using the mean and the standard deviation of scores at k-entry. (Let $M_{i,t}$ denote a score for child i in round $t = 0, 1$ (0 is entry, 1 is exit). Then, the child's standardized score is $\hat{M}_{i,t} = (M_{i,t} - \mu_0)/\sigma_0$, where μ_0 is the estimated sample mean at k-entry, and σ_0 is the estimated standard deviation at k-entry). The deciles of household income in 10,000 2005USD are 1=1.8, 2=3.0, 3=4.1, 4=4.9, 5=6.0, 6=7.2, 7=8.4, 8=10.2, and 9=13.2.

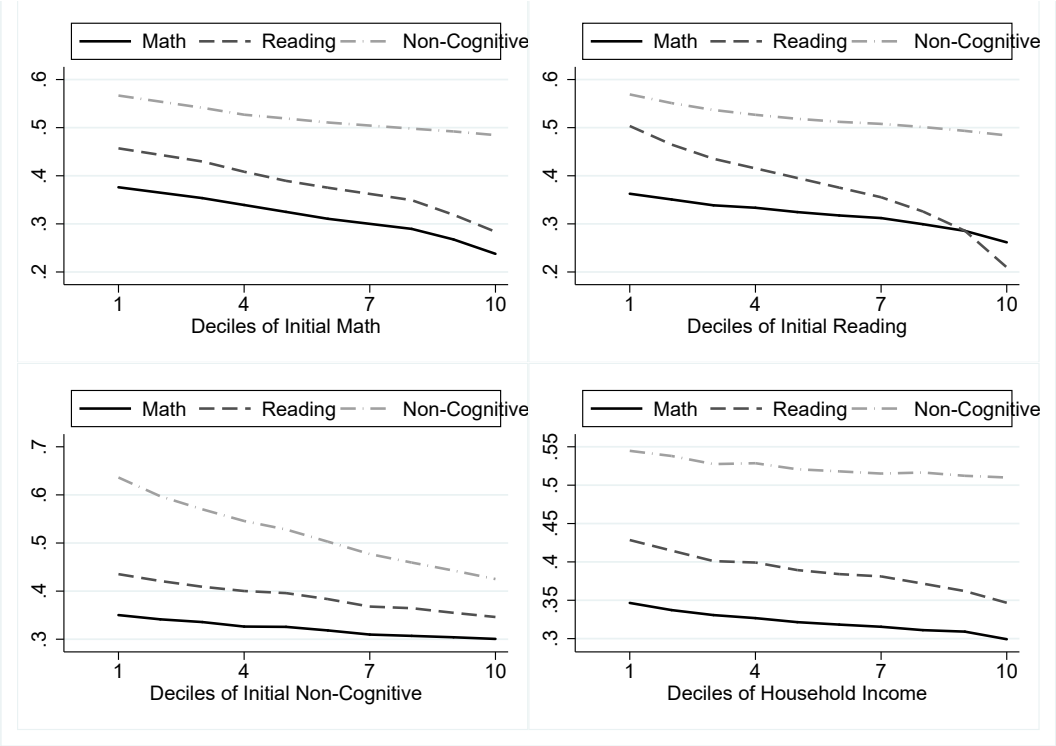
Figure 2: Change in Cognitive and Non-cognitive Scores During Kindergarten by Income Deciles



Source: ECLS-K:1998-99.

Notes: The Math and Reading scores are raw scores. The non-cognitive score is the approaches to learning score evaluated by the teacher. The scores are standardized using the mean and the standard deviation of scores at k-entry. (Let $M_{i,t}$ denote a score for child i in round $t = 0, 1$ (0 is entry, 1 is exit). Then, the child's standardized score is $\hat{M}_{i,t} = (M_{i,t} - \mu_0)/\sigma_0$, where μ_0 is the estimated sample mean at k-entry, and σ_0 is the estimated standard deviation at k-entry). The change is the difference between k-exit and k-entry standardized scores ($\hat{M}_{i,1} - \hat{M}_{i,0}$). The deciles of household income in 10,000 2005USD are 1=1.8, 2=3.0, 3=4.1, 4=4.9, 5=6.0, 6=7.2, 7=8.4, 8=10.2, and 9=13.2.

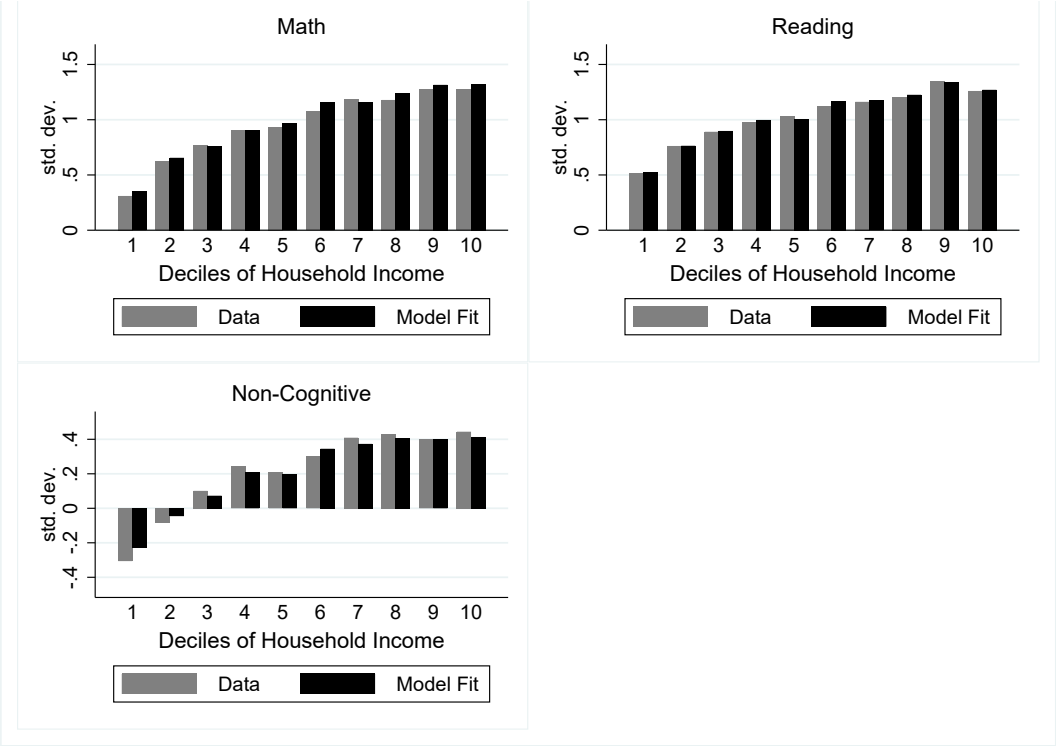
Figure 3: Elasticities of Classroom Investments



Source: ECLS-K:1998-99.

Notes: This figure shows the estimated heterogeneity in the elasticities of skills with respect to classroom investments. For any skill j, the elasticity of children's skills with respect to classroom investments is $\frac{\partial \ln \theta_{j,i,t+1}}{\partial \ln C_{j,i,t}} = \gamma_{2,j} + \gamma_{4,j} \ln \theta_{j,i,t}$. Results pertain to the model with complementarities (see Table 6).

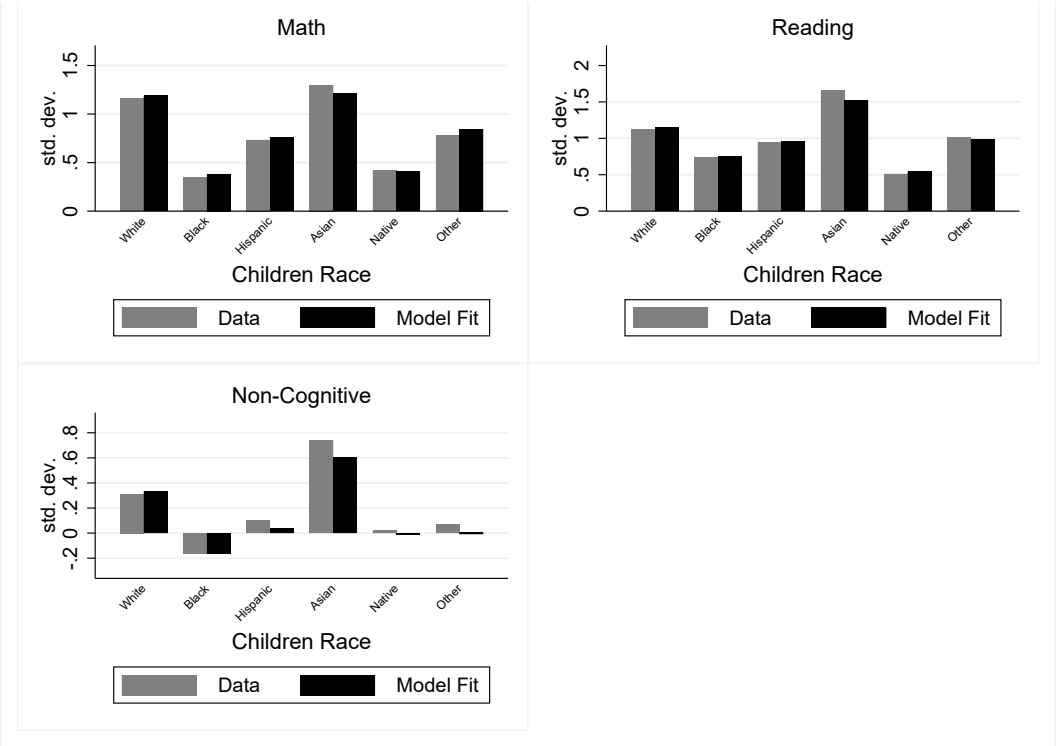
Figure 4: Model Validation: Out-of-Sample Fit of Average Skills by Income



Source: ECLS-K:1998-99.

Notes: This figure displays the results for the validation exercise with respect to family income. Each bar represents the mean children's skills (math, reading or non-cognitive) by income deciles in the data (grey) and predicted by the estimated model (black), respectively. Family income is not an included variable in the estimated model; model predictions for family income are therefore outside of the model.

Figure 5: Model Validation: Out-of-Sample Fit of Average Skill by Child's Race

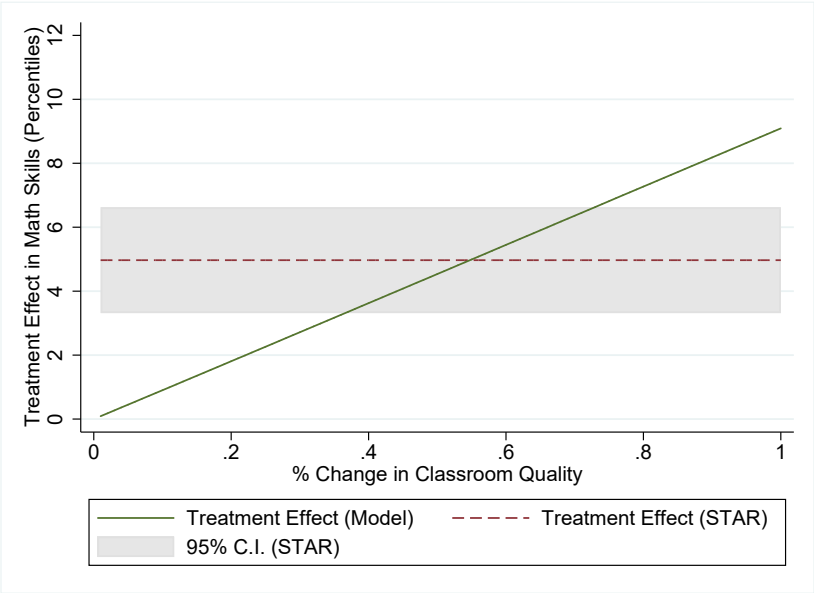


Source: ECLS-K:1998-99.

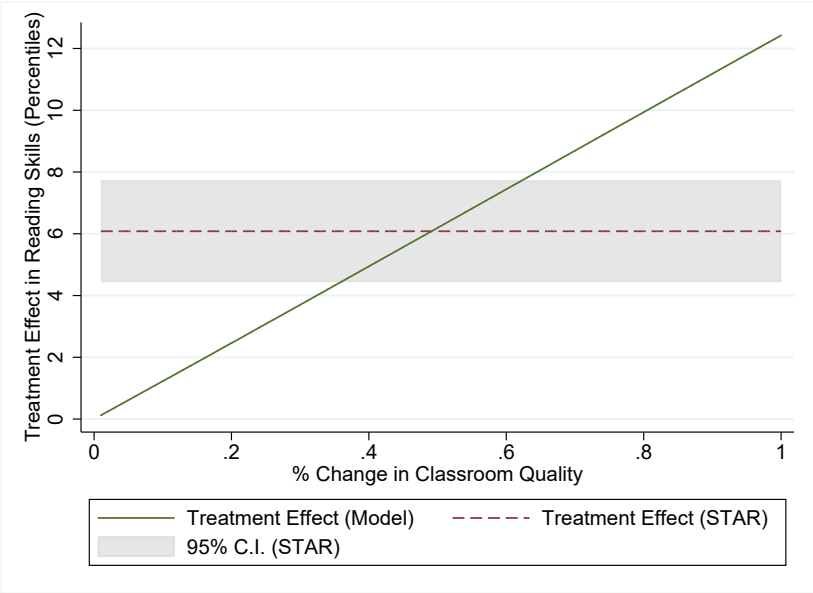
Notes: This table displays the results for the validation exercise with respect to the child's race: White (non-Hispanic), Black, Hispanic, Asian, Native American and Others. Each bar represents the mean children's skills (math, reading or non-cognitive) by the child's race in the data (grey) and predicted by the estimated model (black), respectively. Race is not an included variable in the estimated model; model predictions for race are therefore outside of the model.

Figure 6: Model Validation: STAR and Classroom Quality

Panel A: Math



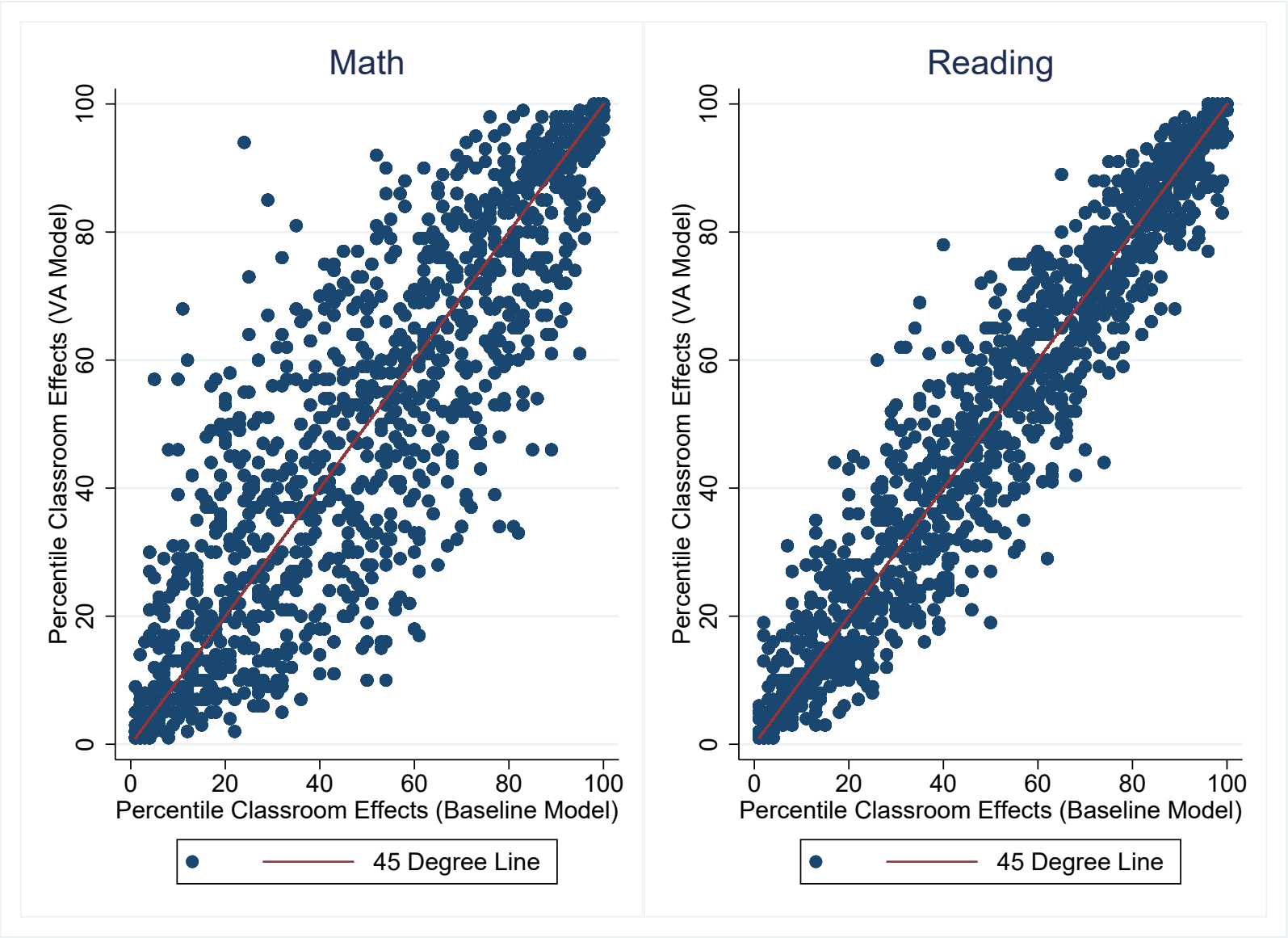
Panel B: Reading



Source: ECLS-K and STAR data.

Notes: This figure shows how the average treatment effect in STAR of reducing classroom size translates into classroom quality. Panel A shows the results for Math Skills, while Panel B shows the results for Reading Skills.

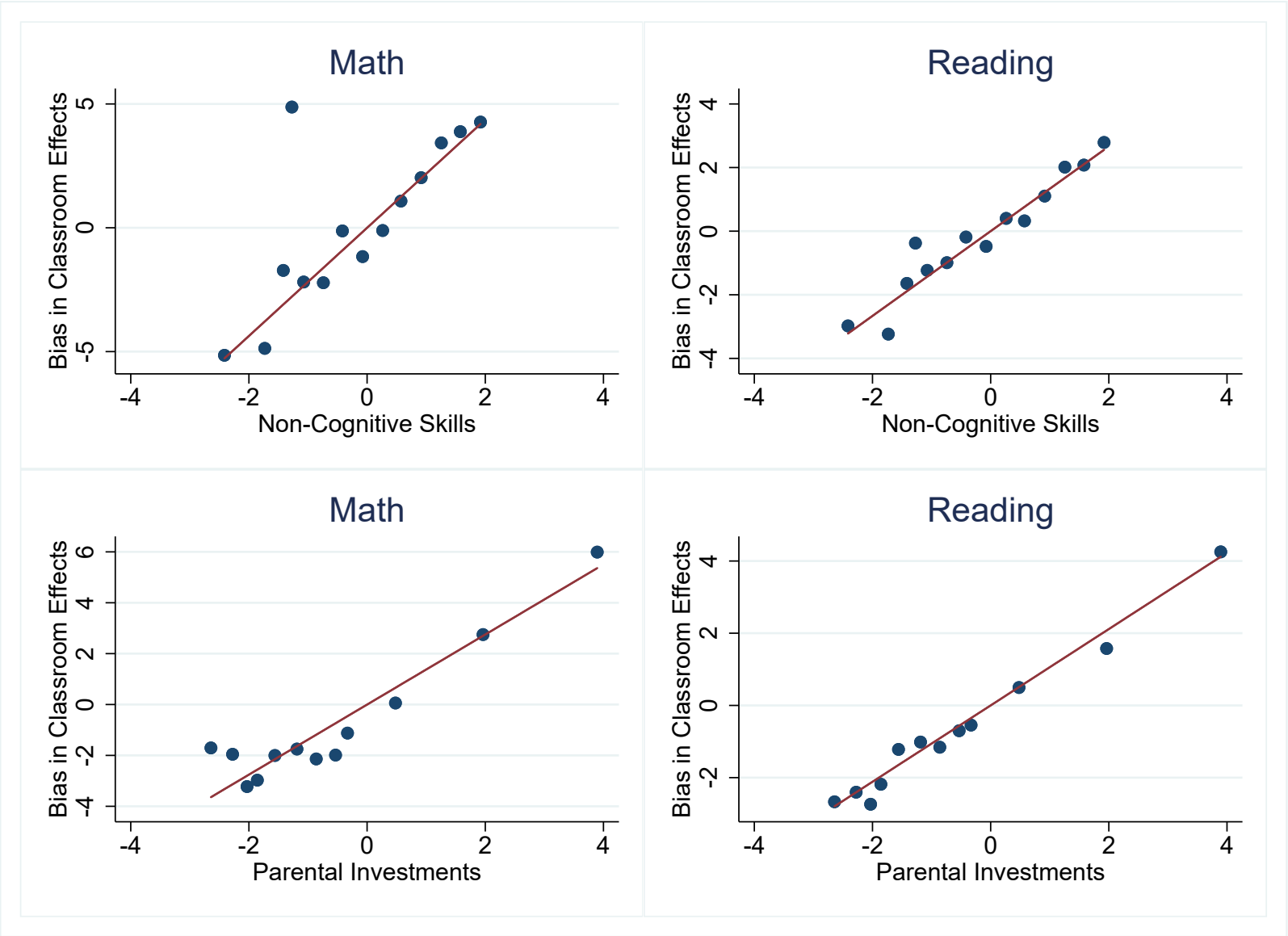
Figure 7: Estimating Classroom Quality: Bias in VA Models



Source: ECLS-K:1998-99

Notes: This figure shows the relationship between the estimated classroom quality in a standard VA linear model versus our unrestricted model. The solid line represents the 45 degree line. The value-added bias is defined as the difference between the rank of estimated latent classroom quality in the standard VA model and our unrestricted model $Bias^{VA} = \widehat{C}^{VA}_{Rank} - \widehat{C}^{UR}_{Rank}$.

Figure 8: Bias in VA Models and Omitted Non-Cognitive and Parental Investments Information



Source:ECLS-K:1998-99

Notes: This figure shows the relationship between the classroom quality bias (difference between VA and baseline estimates) with respect to the average classroom children's non-cognitive skills and parental investments. The value-added bias is defined as the difference between the rank of estimated latent classroom quality in the standard VA model and our unrestricted model $Bias^{VA} = \widehat{C}^{VA}_{Rank} - \widehat{C}^{UR}_{Rank}$. The solid line represents the best linear fit. Average classroom non-cognitive skills and parental investments are divided into 20 quantiles.

Table 1: Descriptive Statistics at Kindergarten Entry

Variable	Mean	Std. Dev.
A: Characteristics of Child		
Number of children	8,656	
Age	5.68	0.36
Fraction male	0.51	
Fraction White, Non-Hispanic	0.68	
Fraction Black, Non-Hispanic	0.14	
Fraction Hispanic	0.09	
Fraction other Race/Ethnicity	0.09	
Fraction living with both biological parents	0.69	
Fraction living with one biological parent	0.27	
Fraction living with no biological parent	0.04	
Fraction having no sibling	0.17	
Fraction having one sibling	0.45	
Fraction having two siblings	0.26	
Fraction having three or more siblings	0.12	
B: Characteristics of Household		
Mother's age	33.89	6.36
Father's age	36.77	6.69
Mother's years of schooling	13.88	2.23
Father's years of Schooling	14.13	2.56
Mother's hours worked	26.13	19.01
Father's hours worked	46.05	13.91
Household income (2017 USD)	68,226	35,480
C: Characteristics of Classroom		
Number of classrooms	1,118	
Class size	20.11	4.65
Instructional time (hours/week)	24.03	9.26
Fraction Teachers Female	0.99	
Teacher's age	42.23	9.79
Teacher years of experience teaching K	9.53	7.88
Fraction of teachers having at least a master's degree	0.35	
D: Characteristics of School		
Number of schools	637	
School year length (days)	178.25	3.13
Fraction public school	0.69	
Fraction of students receiving free or reduced price lunch	0.26	0.27

Source: ECLS-K: 1998-99.

Table 2: Children's Scores and Home Investments at Kindergarten-Entry

	Obs	Mean	SD	Min	Max
Approach to Learning	8656	3.04	0.66	1.00	4.00
Self Control	8656	3.11	0.61	1.00	4.00
Interpersonal Skills	8430	3.01	0.62	1.00	4.00
Math (Routing)	8656	5.11	2.94	0.00	16.00
Math (IRT)	8656	27.56	9.22	10.51	96.04
Reading (Routing)	8656	6.25	3.97	0.00	20.00
Reading (IRT)	8656	36.28	10.45	21.45	138.51
Number of Books	8656	84.77	60.30	0.00	200.00
Computer is Available (Yes/No)	8656	0.63	0.48	0.00	1.00
Mother's Years of Education	8656	13.88	2.23	8.00	20.00

Source: ECLS-K:1998-99.

Notes: The math and reading routing scores are a count of the number of items a child answers correctly on the routing test. The math and reading IRT score are an estimate of the number of items that a child would have answered correctly had she taken all of the questions on all forms. Non-cognitive scores are rated by the teacher on a scale of 1 to 4: 1=Never, 2=Sometimes, 3=Often, 4=Very Often. See the Data Section for more information.

Table 3: Estimates of Measurement Parameters at Kindergarten-Entry

Latent		Location	Scale	Signal to Noise Ratio
Math Skills	Math (Routing)	5.11	2.84	0.93
	Math (IRT)	27.56	8.84	0.92
Reading Skills	Reading (Routing)	6.25	3.94	0.98
	Reading (IRT)	36.28	9.51	0.83
Non-cognitive Skills	Approach to Learning	3.04	0.50	0.58
	Self Control	3.11	0.53	0.76
	Interpersonal Skills	3.01	0.56	0.81
Home Investment	Number of Books	84.77	29.49	0.24
	Computer is Available (Y/N)	0.63	0.25	0.27
	Mother's Education	13.88	1.38	0.38

Source: Model estimates using a sample of ECLS-K data.

Notes: The estimates are for the initial period ($t = 0$). For each measure $M_{\omega,t,m}$ of latent ω at time t, the location is $\mu_{\omega,t,m}$, the scale is $\lambda_{\omega,t,m}$, and the signal-to-noise ratio is $1 - var(\epsilon_{\omega,t,m})/var(M_{\omega,t,m})$. See the measurement equation 4 for more details.

Table 4: Baseline Skill Formation Estimates (Model 1)

	Not Measurement Error Corrected (OLS)			Measurement Error Corrected (IV)		
	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$
$\ln \theta_{M,0}$	0.599	0.215	0.187	0.741	0.225	0.168
	(0.011)	(0.010)	(0.014)	(0.014)	(0.015)	(0.016)
	[0.579,0.620]	[0.193,0.232]	[0.162,0.215]	[0.714,0.767]	[0.195,0.252]	[0.137,0.198]
$\ln \theta_{R,0}$	0.133	0.492	0.082	0.064	0.540	0.016
	(0.012)	(0.012)	(0.012)	(0.015)	(0.014)	(0.014)
	[0.111,0.156]	[0.473,0.515]	[0.056,0.102]	[0.038,0.094]	[0.514,0.569]	[-0.017,0.040]
$\ln \theta_{N,0}$	0.139	0.089	0.606	0.094	0.064	0.738
	(0.009)	(0.009)	(0.012)	(0.013)	(0.014)	(0.018)
	[0.124,0.160]	[0.071,0.106]	[0.584,0.628]	[0.066,0.112]	[0.033,0.089]	[0.695,0.772]
$\ln C_0$	0.337	0.390	0.520	0.323	0.381	0.519
	(0.011)	(0.011)	(0.016)	(0.010)	(0.011)	(0.016)
	[0.320,0.357]	[0.371,0.413]	[0.490,0.546]	[0.305,0.343]	[0.362,0.405]	[0.487,0.548]
N-Children	8656	8656	8656	8656	8656	8656
N-Classroom	1118	1118	1118	1118	1118	1118

Source: ECLS-K:1998-99.
Notes: Classroom-clustered bootstrapped standard errors and 95% confidence intervals are in parentheses and brackets respectively. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table 5: Skill Formation Estimates Including Parental Investments (Model 2)

	Not Measurement Error Corrected (OLS)			Measurement Error Corrected (IV)		
	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$
$\ln \theta_{M,0}$	0.596	0.214	0.185	0.722	0.211	0.157
	(0.011)	(0.010)	(0.014)	(0.015)	(0.016)	(0.017)
	[0.575,0.616]	[0.193,0.232]	[0.160,0.213]	[0.693,0.751]	[0.181,0.238]	[0.125,0.187]
$\ln \theta_{R,0}$	0.131	0.492	0.081	0.044	0.526	0.004
	(0.012)	(0.012)	(0.012)	(0.017)	(0.014)	(0.016)
	[0.108,0.154]	[0.473,0.515]	[0.054,0.102]	[0.015,0.084]	[0.502,0.555]	[-0.036,0.031]
$\ln \theta_{N,0}$	0.138	0.089	0.606	0.078	0.053	0.727
	(0.009)	(0.009)	(0.012)	(0.014)	(0.015)	(0.019)
	[0.123,0.159]	[0.071,0.106]	[0.583,0.628]	[0.048,0.104]	[0.021,0.080]	[0.680,0.760]
$\ln H_0$	0.018	0.003	0.009	0.130	0.094	0.083
	(0.005)	(0.004)	(0.004)	(0.027)	(0.027)	(0.033)
	[0.009,0.027]	[-0.005,0.011]	[0.003,0.018]	[0.081,0.186]	[0.043,0.140]	[0.028,0.152]
$\ln C_0$	0.333	0.391	0.520	0.327	0.396	0.526
	(0.010)	(0.011)	(0.016)	(0.012)	(0.014)	(0.017)
	[0.316,0.354]	[0.371,0.413]	[0.490,0.547]	[0.309,0.356]	[0.374,0.424]	[0.494,0.557]
N-Children	8656	8656	8656	8656	8656	8656
N-Classroom	1118	1118	1118	1118	1118	1118

Source: ECLS-K:1998-99.

Notes: Classroom-clustered bootstrapped standard errors and 95% confidence intervals are in parentheses and brackets respectively. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table 6: Skill Formation Estimates Including Complementarities (Model 3)

	Not Measurement Error Corrected (OLS)			Measurement Error Corrected (IV)		
	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$
$\ln \theta_{M,0}$	0.601	0.212	0.185	0.719	0.211	0.160
	(0.010)	(0.010)	(0.013)	(0.015)	(0.015)	(0.017)
	[0.580,0.619]	[0.189,0.229]	[0.160,0.213]	[0.688,0.750]	[0.178,0.234]	[0.125,0.191]
$\ln \theta_{R,0}$	0.131	0.499	0.082	0.046	0.532	0.006
	(0.012)	(0.012)	(0.012)	(0.016)	(0.013)	(0.015)
	[0.110,0.153]	[0.481,0.520]	[0.056,0.102]	[0.017,0.082]	[0.505,0.563]	[-0.035,0.032]
$\ln \theta_{N,0}$	0.138	0.090	0.603	0.082	0.055	0.715
	(0.009)	(0.009)	(0.012)	(0.014)	(0.015)	(0.018)
	[0.123,0.158]	[0.071,0.107]	[0.581,0.626]	[0.052,0.105]	[0.024,0.081]	[0.671,0.747]
$\ln H_0$	0.017	0.003	0.009	0.113	0.078	0.079
	(0.005)	(0.004)	(0.004)	(0.028)	(0.026)	(0.031)
	[0.008,0.026]	[-0.005,0.010]	[0.003,0.018]	[0.066,0.173]	[0.023,0.120]	[0.032,0.150]
$\ln C_0$	0.334	0.389	0.518	0.323	0.390	0.524
	(0.011)	(0.011)	(0.016)	(0.012)	(0.013)	(0.017)
	[0.316,0.355]	[0.370,0.413]	[0.487,0.544]	[0.305,0.349]	[0.367,0.417]	[0.493,0.558]
$\ln C_0$ x $\ln \theta_{M,0}$	-0.061			-0.041		
	(0.010)			(0.013)		
	[-0.080,-0.038]			[-0.069,-0.022]		
$\ln C_0$ x $\ln \theta_{R,0}$		-0.078			-0.079	
		(0.008)			(0.007)	
		[-0.094,-0.063]			[-0.094,-0.064]	
$\ln C_0$ x $\ln \theta_{N,0}$			-0.061			-0.052
			(0.010)			(0.017)
			[-0.078,-0.042]			[-0.083,-0.019]
N-Children	8656	8656	8656	8656	8656	8656
N-Classroom	1118	1118	1118	1118	1118	1118

Source: ECLS-K:1998-99.

Notes: Classroom-clustered bootstrapped standard errors and 95% confidence intervals are in parentheses and brackets respectively. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table 7: Selection on Observables

Residuals	$\ln \theta_{M,1}$	$\ln \theta_{M,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{R,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{N,1}$	$\ln \theta_{N,1}$
Family Income (100k $)	0.00	0.02	0.02	-0.00	-0.02	-0.02	0.02	0.01	0.01
	[-.029,.032]	[-.03,.06]	[-.03,.06]	[-.03,.02]	[-.06,.02]	[-.06 ,.03]	[-.01,.06]	[-.04,.06]	[-.03,.06]
No Controls	✓			✓			✓		
Family Demographics		✓	✓		✓	✓		✓	✓
Mother and Father Hours Worked			✓			✓			✓

Source: ECLS-K:1998-99.

Notes: Each entry is the estimate from a separate regression. For example, 0.02 in the third column suggests that an increase of $100,000 in family income predicts an increase of 0.02 standard deviations (of skills at kindergarten entry) in the residuals of math skills (not statistically different from zero). Classroom-clustered bootstrapped 95% confidence intervals are in brackets respectively. Family Demographics include: father's education, mother and father's age, household's size and race.

Table 8: Comparison of Model Predictions with STAR

	(1)	(2)	(3)	(4)
Panel A: Comparison of the Latent Classroom Quality				
	Model		STAR	
	Math	Reading	Math	Reading
Classroom Quality Effect (SD)	0.316	0.394	0.361	0.356
Panel B: Comparison of the Heterogeneous Treatment Effects				
	Model		STAR	
	Math	Reading	Math	Reading
Black / White (Ratio)	1.08	1.06	1.08	1.20
			[0.45, 1.62]	[0.65, 1.70]
Free or Reduced-Price Lunch / No Free Lunch (Ratio)	1.09	1.11	0.95	1.13
			[0.46, 1.45]	[0.63, 1.76]

Source: Model estimates using ECLS-K and STAR data.

Notes: Panel A shows the comparison of the estimated classroom quality effect between ECLS-K and STAR data. The outcome in columns (1) and (3) is the math test score, while the outcome in columns (2) and (4) is the reading test score. Columns (1) and (2) show the marginal effects of the ECLS-K classroom latent classroom quality previously estimated from our main model (Table 6). Columns (3) and (4) show the marginal effects of classroom quality estimated from the experimental variation in STAR. Panel B shows the comparison of treatment effect ratios by sub-groups. The class size effect in STAR is estimated by an IV estimation of class size on test scores, using experimental variation in classroom assignment as the instrument. Both white and black children are non-hispanic. In the model, we predict, using the estimated model, the effect of changing classroom quality on children's skills. The model's results are based on the model which includes parental investments and complementarities. Classroom-clustered bootstrapped 95% confidence intervals are in brackets.

Table 9: Value-Added Bias and Teacher's Characteristics

	(1)	(2)	(3)	(4)
	Math		Reading	
Age	0.05	0.02	0.03	0.01
	(0.02)	(0.01)	(0.01)	(0.01)
Black Teacher	-3.02	-5.08	-5.51	-9.90
	(3.93)	(2.82)	(2.26)	(1.78)
Black Teacher X Age	0.12	0.14	0.11	0.25
	(0.09)	(0.06)	(0.05)	(0.04)
Obs	7,918	7,918	7,918	7,918
School F.E.	No	Yes	No	Yes

Source: ECLS-K:1998-99.

Notes: This table shows how the value-added bias is associated with teacher's individual characteristics. The value-added bias is defined as the difference between the rank of estimated latent classroom quality in the standard VA model and our unrestricted baseline model ($Bias^{VA} = \widehat{C}^{VA}_{Rank} - \widehat{C}^{UR}_{Rank}$). Columns (1) and (2) show the results for the bias in math value-added estimates, while columns (3) and (4) show the results for the bias in reading value-added estimates. The even columns report results including school fixed effects.

Table 10: Initial Conditions

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\ln \theta_{M,0}$	(1)	1.00						
$\ln \theta_{R,0}$	(2)	0.69	1.00					
$\ln \theta_{N,0}$	(3)	0.53	0.47	1.00				
$\ln H_0$	(4)	0.53	0.42	0.33	1.00			
$\ln C_{M,0}$	(5)	0.04	0.07	-0.09	-0.15	1.00		
$\ln C_{R,0}$	(6)	0.03	0.09	-0.07	-0.22	0.49	1.00	
$\ln C_{N,0}$	(7)	-0.06	-0.06	-0.08	-0.17	0.08	0.10	1.00

Source: Model estimates using a sample of ECLS data.

Notes: Results pertain to the model with complementarities.

Table 11: Income Gradient in Classroom Quality and Parental Investments

	$\ln C_{M,0}$	$\ln C_{R,0}$	$\ln C_{N,0}$	$\ln H_0$
Below Median x Income ($10k)	0.055	0.040	-0.000	0.302
	(0.016)	(0.016)	(0.016)	(0.031)
Above Median x Income ($10k)	-0.010	-0.010	-0.006	0.095
	(0.013)	(0.013)	(0.013)	(0.025)
N	5789	5789	5789	5789

Source: Model estimates using a sample of ECLS-K data.

Notes: Each column shows the estimates from a separate regression of classroom quality (math, reading or non-cognitive) on family income interacted with a dummy for whether family income is above or below the median of the sample distribution. Median family income in the sample is $45,000.

Table 12: Decomposition of School and Family

Variable	(1) $E[\theta \mid Income = 10^{th}]$	(2) $E[\theta \mid Income = 90^{th}]$	(3) $E[\theta \mid Income = 10^{th}]$ $-E[\theta \mid Income = 90^{th}]$	(4) $E[\theta \mid Income = 10^{th}, C = 90^{th}]$ $-E[\theta \mid Income = 90^{th}]$	(5) $E[\theta \mid Income = 10^{th}, H = 90^{th}]$ $-E[\theta \mid Income = 90^{th}]$
Panel A: Measurement Error Correction and Complementarities					
Math	0.40	1.28	-0.88	-0.81 (-7.95%)	-0.72 (-18.18%)
Reading	0.62	1.31	-0.69	-0.66 (-4.35%)	-0.59 (-14.49%)
Non-Cognitive	-0.24	0.42	-0.66	-0.60 (-9.09%)	-0.50 (-24.24%)
Panel B: OLS with Linear Model					
Math	0.40	1.28	-0.88	-0.69 (-21.59%)	-0.80 (-7.22%)
Reading	0.62	1.31	-0.69	-0.57 (-17.39%)	-0.65 (-5.80%)
Non-Cognitive	-0.24	0.42	-0.66	-0.51 (-22.73%)	-0.53 (-19.70%)
Panel C: Simple VA Model					
Math	0.40	1.28	-0.88	-0.77 (-12.5%)	
Reading	0.62	1.31	-0.69	-0.64 (-7.25%)	

Source: Model estimates using a sample of ECLS-K data.

Notes: This table shows the Achievement Gap Decomposition (relative to the 10-90 deciles of Income Distribution) with respect to school and home investments. Panel A displays the results for the measurement error corrected model which allows for complementarities. Panel B shows the results for a linear model without measurement error correction. The first two columns represents the average of skills by income deciles (10th and 90th). The third column shows the gap between these two subpopulations (differences between column (1) and (2)). Column (4) shows the skills gap once the children from the low-income subgroup are matched with the average school quality from the high-income children (and the percentage change with respect to column (3) in parenthesis). Column (5) shows the skills gap once the children from the low-income subgroup are matched with the average parental investments from the high-income children (and the percentage change with respect to column (3) in parenthesis).

Appendices

A Mathematical Derivations

A.1 Identification of Measurement System

The identification of the location parameter is the same for all the measures for every type of skills. It is based on the normalization of the location of the child's skills during the first period:

$$E[M_{\omega,0,m}] = \mu_{\omega,0,m} \text{ for all } \omega_{i,0} \in \{\theta_{M,i,0}, \theta_{R,i,0}, \theta_{N,i,0}\} \tag{A-1}$$

where $\{\theta_{M,i,0}, \theta_{R,i,0}, \theta_{N,i,0}\}$ represents a child's math, reading and non-cognitive skills, respectively.

The identification of the factor loading requires few more algebraic steps. Let's start with the non-cognitive skills, the case in which we have three different measures and we identify the factor loading as:

$$\lambda_{N,0,m} = \sqrt{\frac{Cov(M_{N,0,m}, M_{N,0,m'})Cov(M_{N,0,m}, M_{N,0,m''})}{Cov(M_{N,0,m'}, M_{N,0,m''})}} \; \forall m \tag{A-2}$$

where (m, m', m'') are three different measures for the child's non-cognitive skills.

We now identify the rest of the loading factors using only two different measures for each type of skill. In the case of math skills θ_M we have

$$\sqrt{\frac{Cov(M_{M,0,m}, M_{M,0,m'})Cov(M_{M,0,m}, M_{N,0,m})}{Cov(M_{M,0,m'}, M_{N,0,m})}} =$$
$$\sqrt{\frac{\lambda_{M,0,m} \cdot \lambda_{M,0,m'} \cdot \underbrace{Var(\theta_M)}_{=1} \cdot \lambda_{M,0,m} \cdot \lambda_{N,0,m} \cdot Cov(\theta_M, \theta_N)}{\lambda_{M,0,m'} \cdot \lambda_{N,0,m} \cdot Cov(\theta_M, \theta_N)}} =$$
$$\sqrt{\lambda^2_{M,0,m}} = \lambda_{M,0,m} \;\; \forall m \; ,$$

which shows that we only need two different measures (m, m') for math to identify its measurement parameters once we have at least two skills. Identification of measurement parameters for reading skills (θ_R) is similar.

A.2 Equivalence in Measures Across Datasets

Consider the following measurement system for each dataset:

$$M^{ECLS-K}_{j,i,t,m} = \mu^{ECLS-K}_{j,t,m} + \lambda^{ECLS-K}_{j,t,m} \ln \theta_{j,i,t} + \epsilon^{ECLS-K}_{j,i,t,m} \tag{A-3}$$

$$M^{STAR}_{j,i,t,m'} = \mu^{STAR}_{j,t,m'} + \lambda^{STAR}_{j,t,m'} \ln \theta_{j,i,t} + \epsilon^{STAR}_{j,i,t,m'} \tag{A-4}$$

where the location and scale of the ECLS-K and STAR test scores are unknown, and can be different from each other. The average treatment effect for some treatment τ is the expected difference in test scores between the treatment group (τ=1) and the control group (τ=0):

$$\begin{aligned} E\left[TE^{STAR}\right] &= E\left[M^{STAR}_{j,i,1,m'}|\tau=1\right] - E\left[M^{STAR}_{j,i,1,m'}|\tau=0\right] \\ &= \lambda^{STAR}_{j,1,m'}\Big(E\left[\ln\theta_{j,i,1}|\tau=1\right] - E\left[\ln\theta_{j,i,1}|\tau=0\right]\Big) \end{aligned} \tag{A-5}$$

Note that the treatment effect depends on the scale of the test score, $\lambda^{STAR}_{j,t,m'}$. This creates a problem in comparing average treatment effects between STAR and ECLS-K if considering raw scores. Standardizing test scores does not help because it would just change the scale (λ) by a different standard deviation for each test. However, the relative average treatment effect between two subpopulations is free from the loading factor and can be compared between the two datasets. To see this, consider the binary variable X, which takes a value of 1, for example, if the child qualifies for free or reduced-priced lunch, and 0 otherwise. We define the relative average treatment effect for these two subpopulations as

$$\begin{aligned} \frac{E\left[TE^{STAR}|X=1\right]}{E\left[TE^{STAR}|X=0\right]} &= \frac{E\left[M^{STAR}_{j,i,1,m'}|X=1,\,\tau=1\right] - E\left[M^{STAR}_{j,i,1,m'}|X=1,\,\tau=0\right]}{E\left[M^{STAR}_{j,i,1,m'}|X=0,\,\tau=1\right] - E\left[M^{STAR}_{j,i,1,m'}|X=0,\,\tau=0\right]} \\ &= \frac{E\left[\ln\theta_{j,i,1}|X=1,\,\tau=1\right] - E\left[\ln\theta_{j,i,1}|X=1,\,\tau=0\right]}{E\left[\ln\theta_{j,i,1}|X=0,\,\tau=1\right] - E\left[\ln\theta_{j,i,1}|X=0,\,\tau=0\right]} \end{aligned} \tag{A-6}$$

which is independent of the loading factor.

B Monte Carlo

This section uses Monte Carlo exercises to examine the properties of the estimator. The true data generating process is assumed to be:

$$\ln \theta_{i,1} = \gamma_0 + \gamma_1 \ \ln \theta_{i,0} + \gamma_2 \ \ln C_{i,0} + \gamma_3 \ \ln \theta_{i,0} \ln C_{i,0} + \eta_{i,0} \tag{B-1}$$
$$\ln C_{i,0} \sim N(0,1) \tag{B-2}$$
$$\ln \theta_{i,0} \sim N(1,1) \tag{B-3}$$
$$\eta_{i,0} \sim N(0,0.2) \tag{B-4}$$

where $\theta_{i,0}$ is the skill of child i at time 0, $\ln C_{i,0}$ is the investment from school, and $\eta_{i,0}$ is the error term. We assume that children are clustered by classrooms and we observe multiple children per classroom. Further, we assume that investment from school is the same for all children in the same classroom. We assume that $\ln \theta_{i,0}$ is not observed, but we observe two noisy measures of it:

$$\ln M_{i,0,1} = \ln \theta_{i,0} + \epsilon_{i,0,1} \tag{B-5}$$
$$\ln M_{i,0,2} = \ln \theta_{i,0} + \epsilon_{i,0,2} \tag{B-6}$$
$$\epsilon_{i,0,1} \sim N(0,0.3) \tag{B-7}$$
$$\epsilon_{i,0,2} \sim N(0,0.3) \tag{B-8}$$

We simulate 200 different classrooms for different numbers of children per classroom (n). We generate 100 simulated dataset of from the data generating process described above. Table B-1 reports the mean estimates and their 90% confidence intervals.

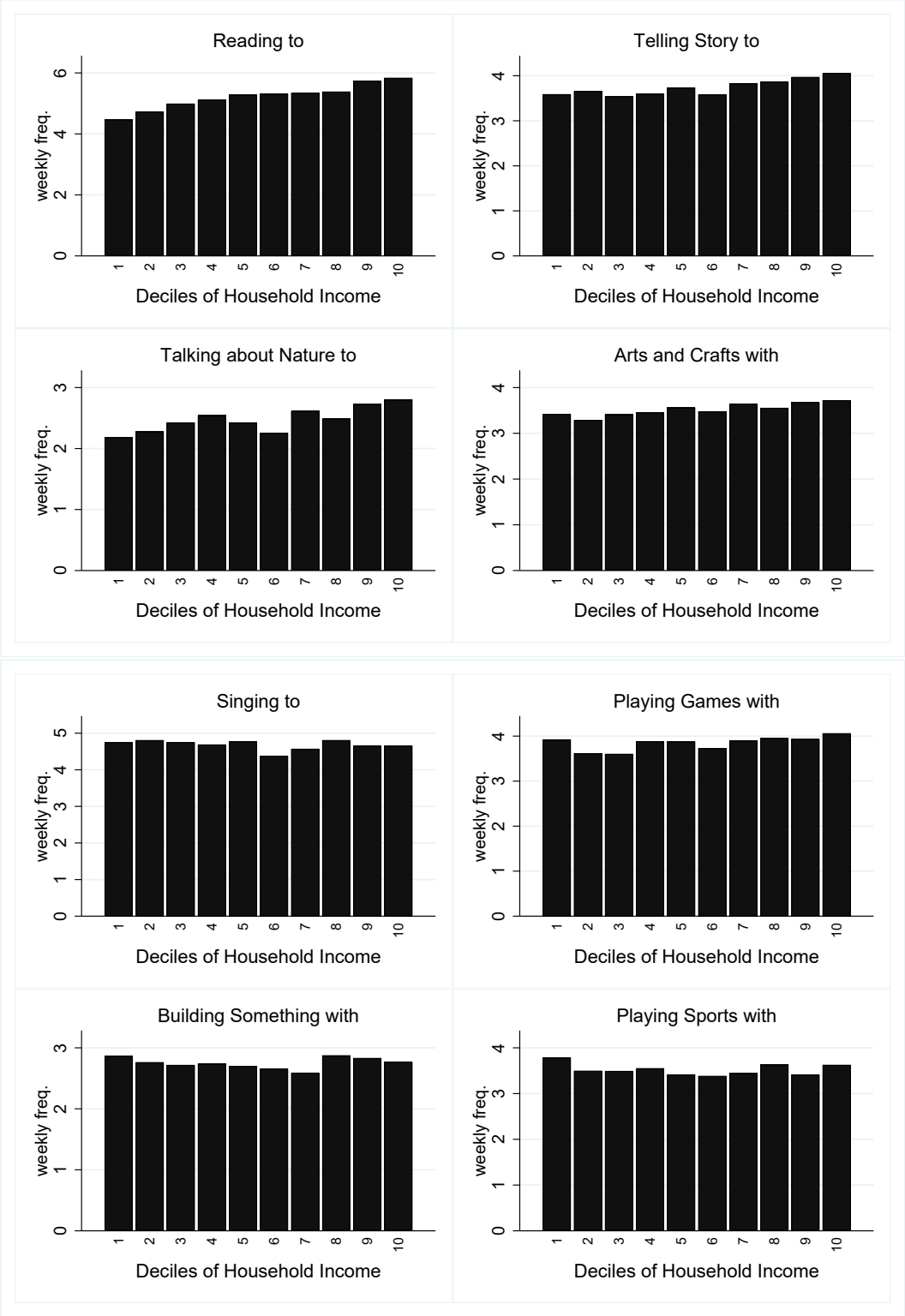
Table B-1: Monte Carlo Results

True Values	γ_0	γ_1	γ_2	γ_3
	4.000	1.000	2.000	3.000
Model Estimates				
$n = 10$	4.001	0.998	1.993	3.008
	[3.958,4.040]	[0.945,1.053]	[1.939,2.042]	[2.954,3.067]
$n = 20$	3.999	0.999	1.999	3.004
	[3.971,4.035]	[0.955,1.038]	[1.952,2.037]	[2.961,3.050]

Source: Monte Carlo results.
Notes: n denotes the number of children per classroom. 90% confidence intervals are in parentheses.

C Supplementary Analysis

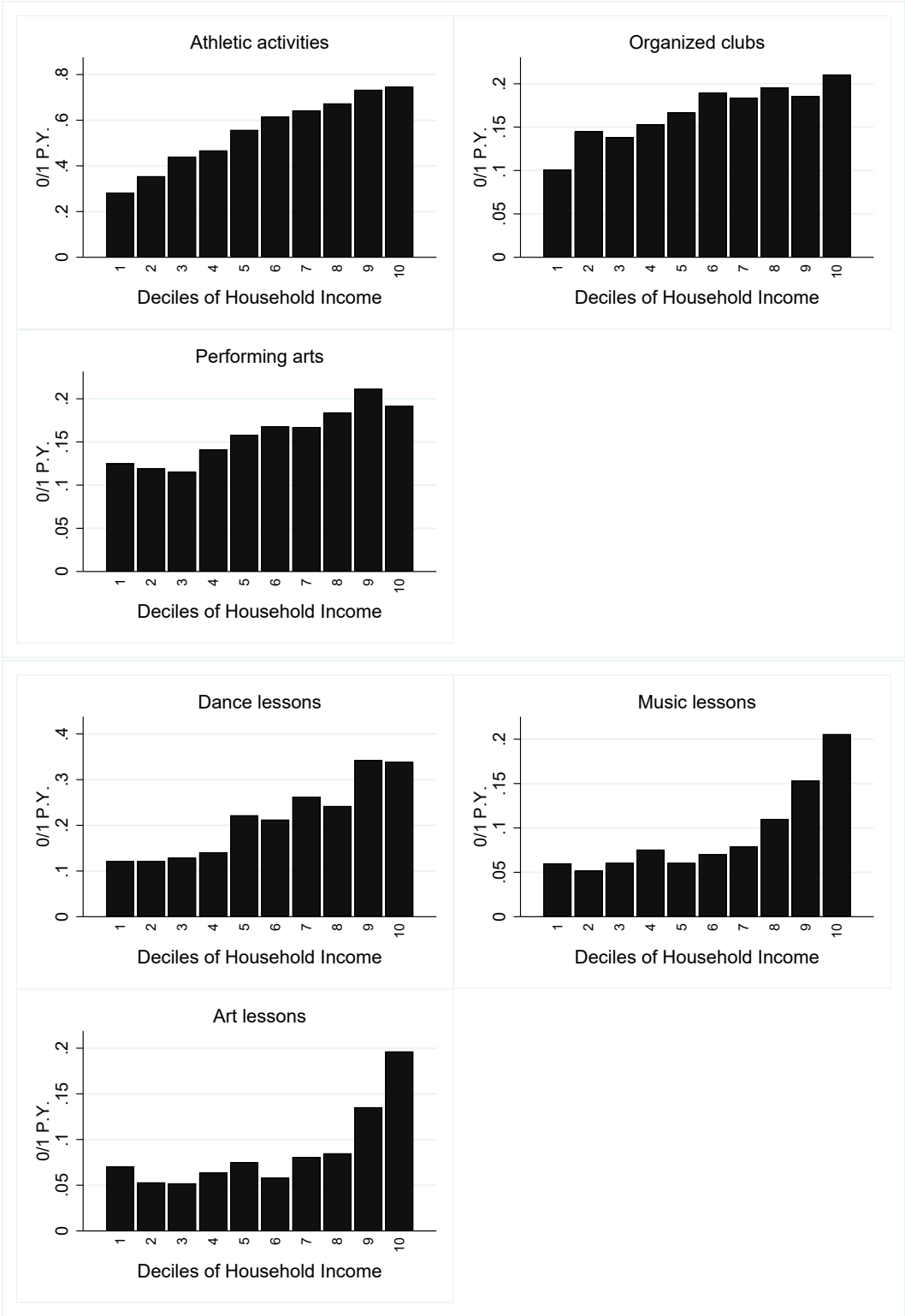
Figure C-1: Home Investments in the Year of Kindergarten over Income Deciles



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data.

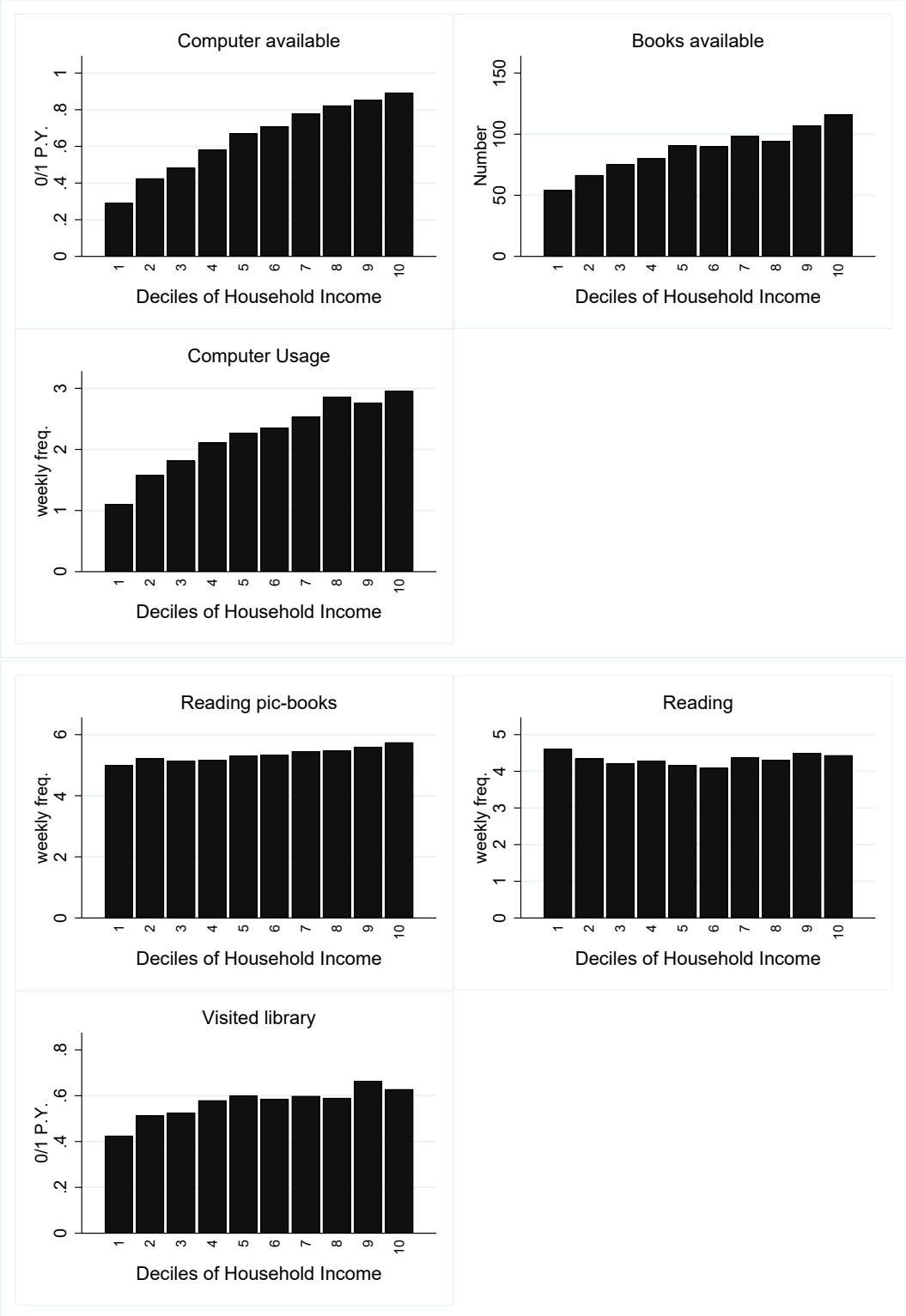
Figure C-2: Home Investments in the Year of Kindergarten over Income Deciles



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data.

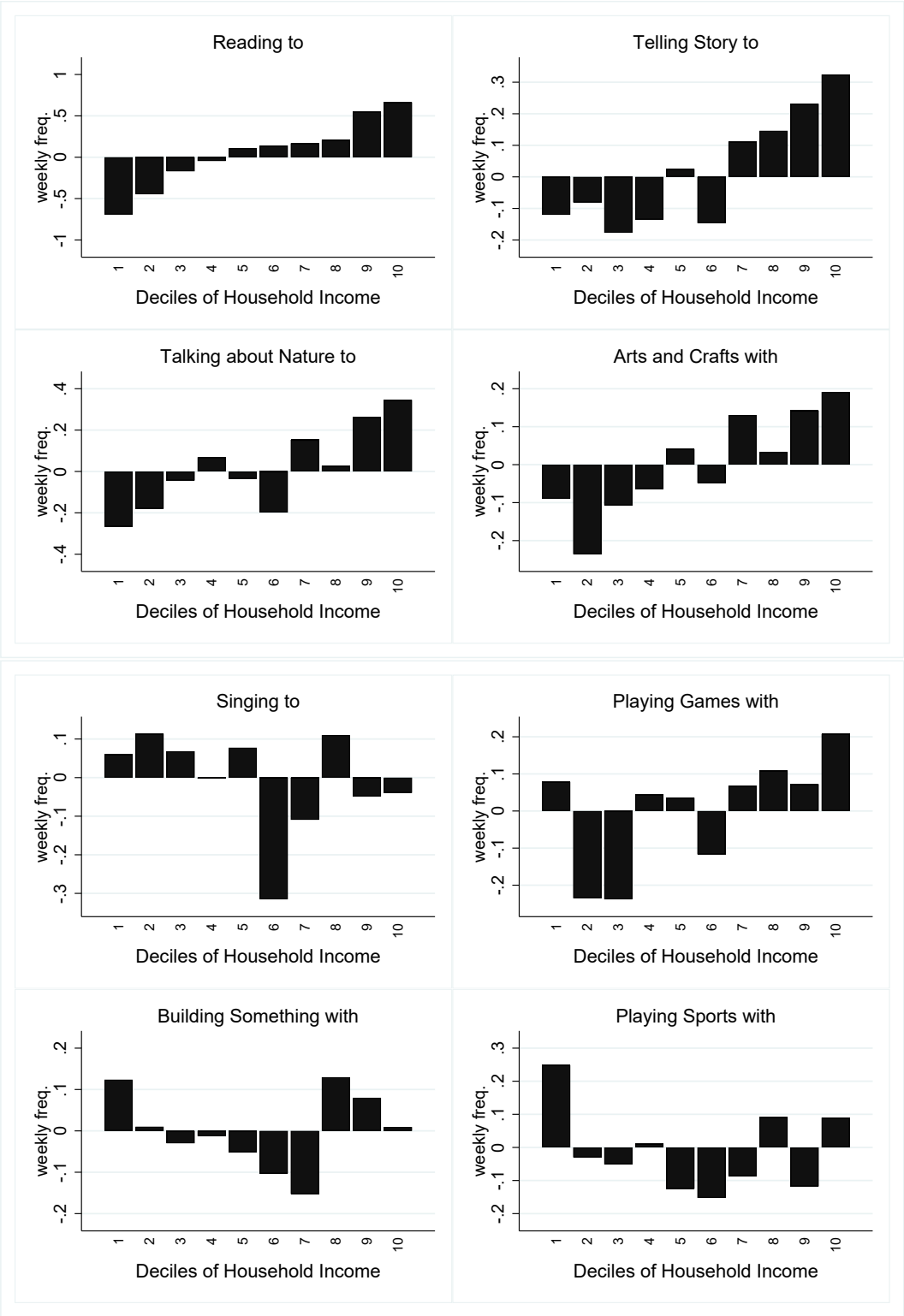
Figure C-3: Home Investments in the Year of Kindergarten over Income Deciles



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data.

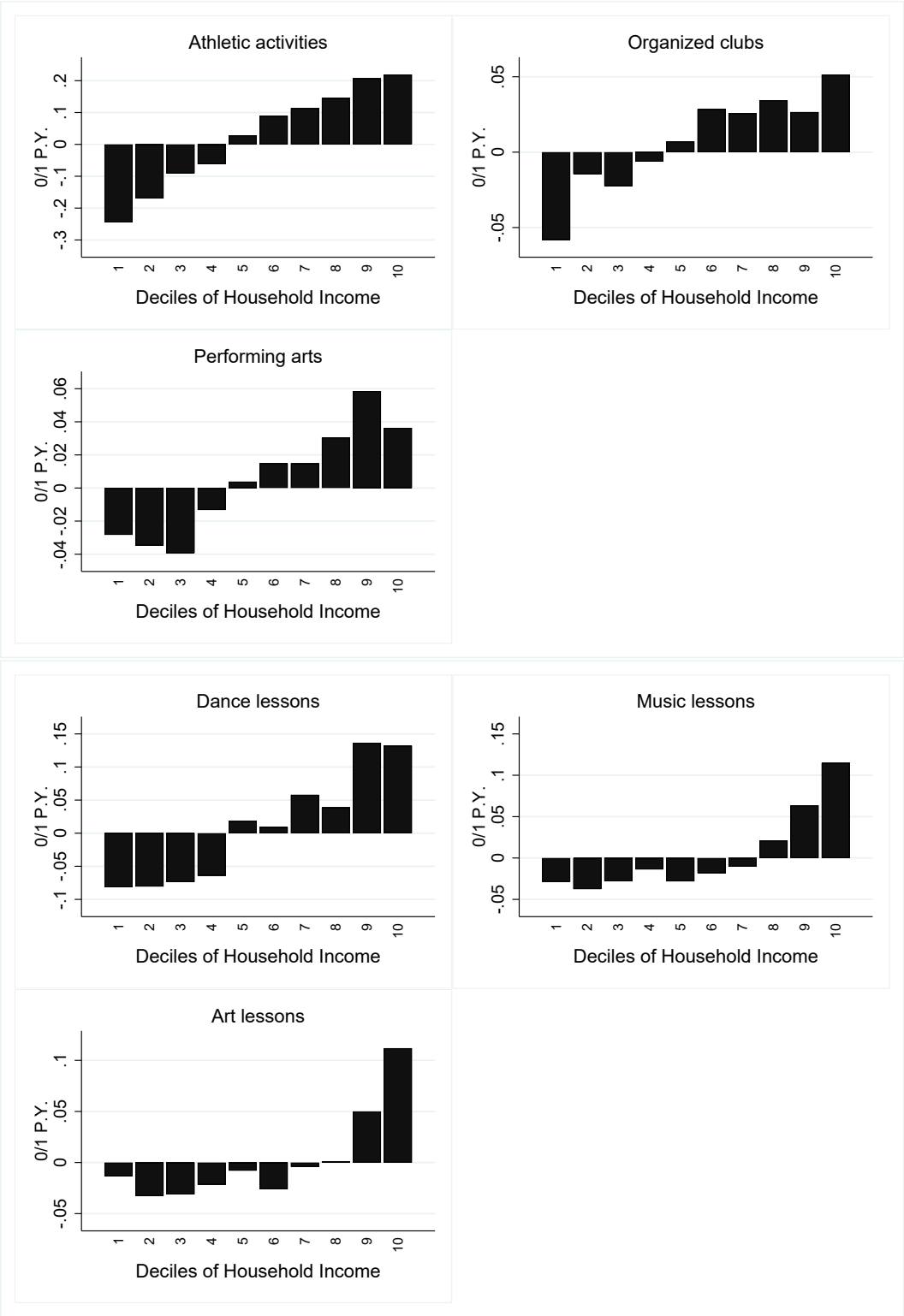
Figure C-4: Home Investments in the Year of Kindergarten over Income Deciles: Age Adjusted



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data. Each parental investment measure is age adjusted, i.e., it is the residual variation after controlling for children's age (in months).

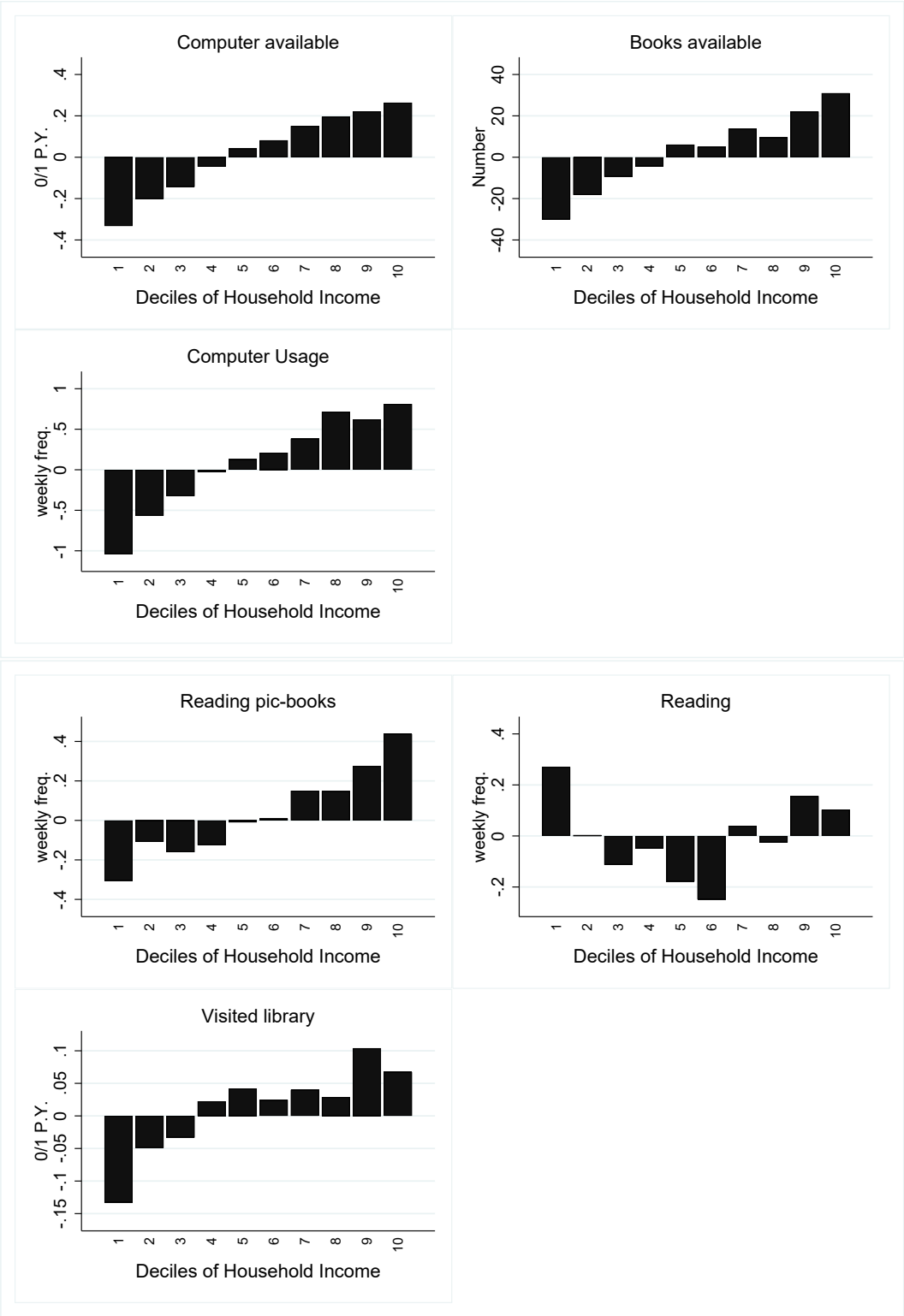
Figure C-5: Home Investments in the Year of Kindergarten over Income Deciles: Age Adjusted



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data. Each parental investment measure is age adjusted, i.e., it is the residual variation after controlling for children's age (in months).

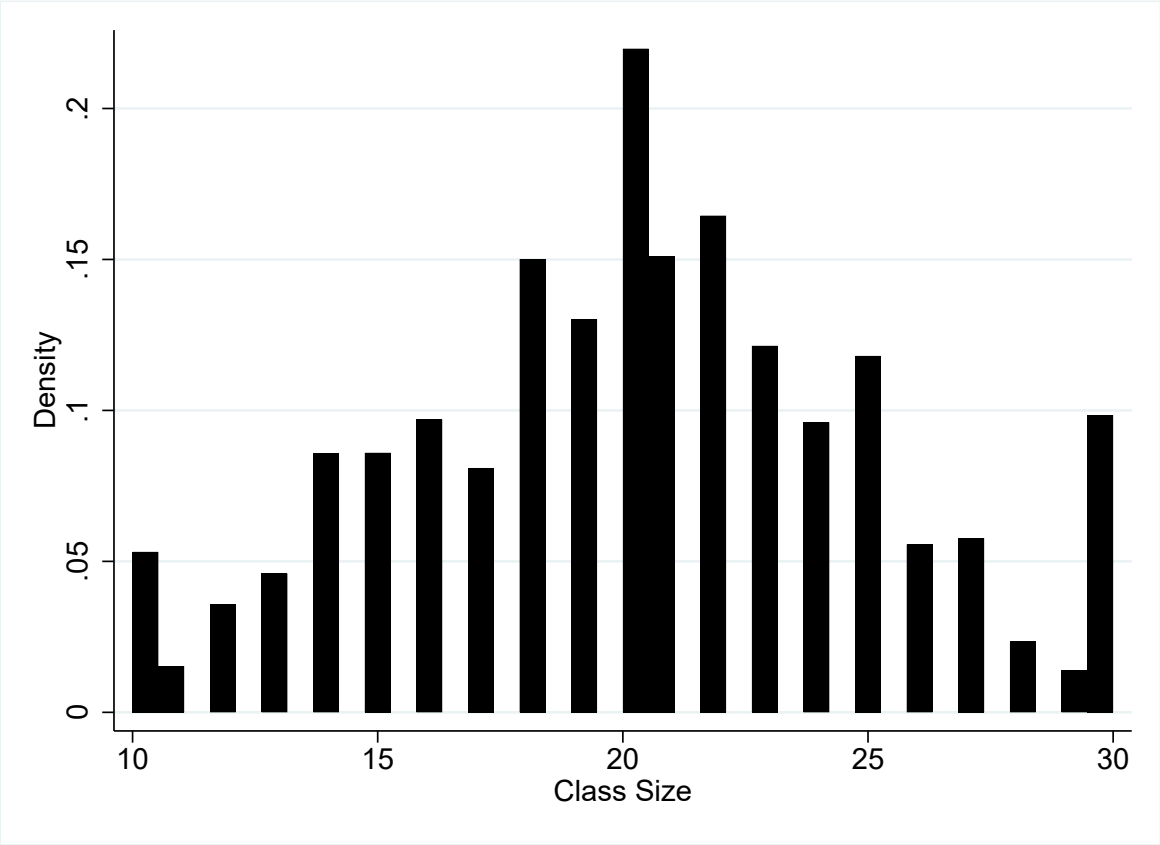
Figure C-6: Home Investments in the Year of Kindergarten over Income Deciles: Age Adjusted



Source: ECLS-K:1998-99.

Notes: This figure shows the distribution of various parental investment measures over the household income deciles in our data. Each parental investment measure is age adjusted, i.e., it is the residual variation after controlling for children's age (in months).

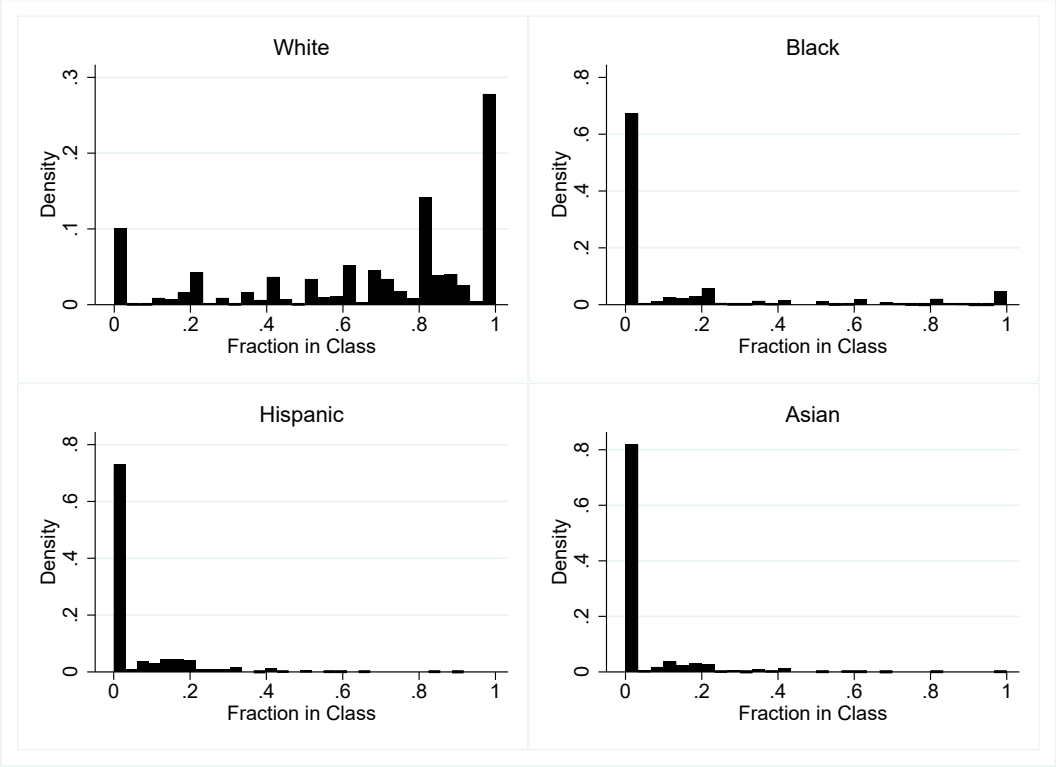
Figure C-7: The Distribution of Class Size



Source: ECLS-K:1998-99.

Notes: This figure shows the empirical distribution of classroom size in our estimating sample.

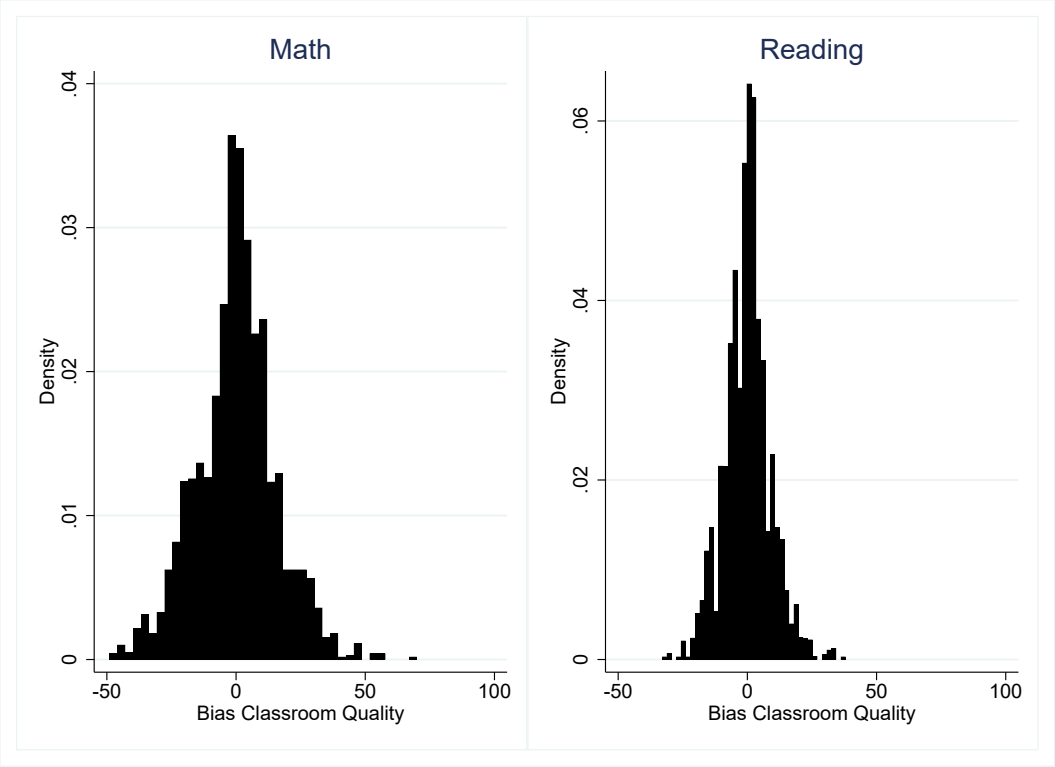
Figure C-8: Distribution of Racial Composition between Classrooms



Source: ECLS-K:1998-99.

Notes: This figure shows the empirical distribution of racail composition between classrooms in our sample.

Figure C-9: Distribution of Bias in Classroom Quality Estimates



Source: ECLS-K:1998-99.

Notes: This figure shows the empirical distribution of bias in classroom quality estimates.

Table C-1: Descriptive Statistics at Kindergarten Entry; Full ECLS-K Data

Variable	Mean	Std. Dev.
A: Characteristics of Child		
Number of children	21,409	
Age	5.66	0.37
Fraction male	0.51	
Fraction White, Non-Hispanic	0.55	
Fraction Black, Non-Hispanic	0.15	
Fraction Hispanic	0.18	
Fraction other Race/Ethnicity	0.12	
Fraction living with both biological parents	0.66	
Fraction living with one biological parent	0.30	
Fraction living with no biological parent	0.04	
Fraction having no sibling	0.18	
Fraction having one sibling	0.42	
Fraction having two siblings	0.26	
Fraction having three or more siblings	0.14	
B: Characteristics of Household		
Mother's age	33.21	6.63
Father's age	36.23	6.99
Mother's years of schooling	13.36	2.38
Father's years of Schooling	13.64	2.69
Mother's hours worked	25.01	19.50
Father's hours worked	44.81	14.25
Household income (2017 USD)	62,432	35,820
C: Characteristics of Classroom		
Number of classrooms	5,224	
Class size	20.27	4.41
Instructional time (hours/week)	23.90	8.91
Fraction Teachers Female	0.98	
Teacher's age	41.34	10.12
Teacher years of experience teaching K	8.76	7.70
Fraction of teachers having at least a master's degree	0.37	
D: Characteristics of School		
Number of schools	1,591	
School year length (days)	178.31	3.14
Fraction public school	0.79	
Fraction of students receiving free or reduced price lunch	0.31	0.28

Source: ECLS-K: 1998-99.

Table C-2: Skill Formation Estimates: Full Interacted Model

	Not Measurement Error Corrected (OLS)			Measurement Error Corrected (IV)		
	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$
$\ln \theta_{M,0}$	0.599	0.212	0.184	0.718	0.206	0.158
	(0.011)	(0.010)	(0.013)	(0.015)	(0.015)	(0.017)
	[0.578,0.619]	[0.189,0.229]	[0.160,0.212]	[0.689,0.747]	[0.174,0.233]	[0.123,0.188]
$\ln \theta_{R,0}$	0.133	0.498	0.080	0.049	0.535	0.004
	(0.012)	(0.012)	(0.012)	(0.016)	(0.013)	(0.015)
	[0.114,0.154]	[0.481,0.519]	[0.055,0.101]	[0.020,0.084]	[0.509,0.563]	[-0.033,0.029]
$\ln \theta_{N,0}$	0.138	0.090	0.604	0.082	0.056	0.716
	(0.009)	(0.009)	(0.012)	(0.013)	(0.015)	(0.018)
	[0.122,0.158]	[0.072,0.107]	[0.581,0.626]	[0.053,0.105]	[0.025,0.084]	[0.674,0.747]
$\ln H_0$	0.017	0.002	0.009	0.111	0.077	0.081
	(0.005)	(0.004)	(0.004)	(0.028)	(0.026)	(0.030)
	[0.008,0.026]	[-0.005,0.010]	[0.003,0.018]	[0.064,0.171]	[0.023,0.120]	[0.034,0.151]
$\ln C_0$	0.335	0.392	0.520	0.323	0.391	0.526
	(0.011)	(0.011)	(0.017)	(0.012)	(0.013)	(0.017)
	[0.317,0.356]	[0.373,0.416]	[0.489,0.545]	[0.306,0.349]	[0.371,0.416]	[0.497,0.554]
$\ln C_0$x $\ln \theta_{M,0}$	-0.042	-0.026	-0.018	-0.024	-0.014	-0.028
	(0.013)	(0.008)	(0.012)	(0.021)	(0.013)	(0.016)
	[-0.067,-0.016]	[-0.044,-0.011]	[-0.046,0.002]	[-0.068,0.016]	[-0.036,0.012]	[-0.052,0.006]
$\ln C_0$x $\ln \theta_{R,0}$	-0.033	-0.069	-0.027	-0.032	-0.074	-0.015
	(0.011)	(0.008)	(0.012)	(0.017)	(0.008)	(0.012)
	[-0.054,-0.014]	[-0.083,-0.053]	[-0.047,-0.003]	[-0.060,0.006]	[-0.092,-0.059]	[-0.038,0.013]
$\ln C_0$x $\ln \theta_{N,0}$	-0.000	0.012	-0.048	0.007	0.012	-0.038
	(0.007)	(0.007)	(0.011)	(0.013)	(0.012)	(0.018)
	[-0.016,0.014]	[-0.005,0.024]	[-0.069,-0.030]	[-0.017,0.034]	[-0.018,0.035]	[-0.078,-0.005]
N-Children	8656	8656	8656	8656	8656	8656
N-Classroom	1118	1118	1118	1118	1118	1118

Source: ECLS-K:1998-99.

Notes: Classroom-clustered bootstrapped standard errors and 95% confidence intervals are in parentheses and brackets respectively. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table C-3: Skill Formation Estimates: Full Interacted Model

	Not Measurement Error Corrected (OLS)			Measurement Error Corrected (IV)		
	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$	$\ln \theta_{M,1}$	$\ln \theta_{R,1}$	$\ln \theta_{N,1}$
$\ln \theta_{M,0}$	0.603	0.212	0.185	0.744	0.205	0.166
	(0.011)	(0.010)	(0.013)	(0.017)	(0.016)	(0.017)
	[0.581,0.623]	[0.189,0.229]	[0.161,0.214]	[0.718,0.781]	[0.171,0.235]	[0.128,0.195]
$\ln \theta_{R,0}$	0.132	0.500	0.083	0.064	0.567	0.014
	(0.012)	(0.012)	(0.012)	(0.015)	(0.022)	(0.015)
	[0.111,0.154]	[0.480,0.522]	[0.056,0.103]	[0.044,0.095]	[0.535,0.627]	[-0.018,0.041]
$\ln \theta_{N,0}$	0.137	0.089	0.601	0.074	0.051	0.707
	(0.009)	(0.009)	(0.012)	(0.014)	(0.015)	(0.019)
	[0.121,0.157]	[0.070,0.106]	[0.579,0.623]	[0.046,0.098]	[0.015,0.075]	[0.674,0.745]
$\ln H_0$	0.018	0.003	0.010	0.100	0.061	0.067
	(0.005)	(0.004)	(0.004)	(0.029)	(0.026)	(0.030)
	[0.009,0.027]	[-0.005,0.011]	[0.003,0.019]	[0.045,0.153]	[0.016,0.110]	[0.021,0.125]
$\ln C_0$	0.333	0.389	0.518	0.321	0.383	0.527
	(0.011)	(0.011)	(0.016)	(0.094)	(0.040)	(0.109)
	[0.315,0.354]	[0.370,0.413]	[0.488,0.544]	[0.004,0.370]	[0.361,0.411]	[0.496,0.562]
$\ln H_0$ x $\ln C_0$	0.001	-0.006	-0.004	-0.032	-0.002	-0.020
	(0.004)	(0.004)	(0.004)	(0.044)	(0.027)	(0.030)
	[-0.007,0.010]	[-0.014,0.003]	[-0.011,0.003]	[-0.104,0.049]	[-0.065,0.036]	[-0.083,0.027]
$\ln C_0$ x $\ln \theta_{M,0}$	-0.060			-0.037		
	(0.010)			(0.025)		
	[-0.080,-0.038]			[-0.083,0.008]		
$\ln C_0$ x $\ln \theta_{R,0}$		-0.077			-0.083	
		(0.008)			(0.014)	
		[-0.092,-0.061]			[-0.103,-0.056]	
$\ln C_0$ x $\ln \theta_{N,0}$			-0.060			-0.056
			(0.010)			(0.026)
			[-0.079,-0.041]			[-0.104,-0.021]
$\ln H_0$ x $\ln \theta_{M,0}$	-0.008			-0.060		
	(0.003)			(0.016)		
	[-0.015,-0.002]			[-0.098,-0.032]		
$\ln H_0$ x $\ln \theta_{R,0}$		-0.006			-0.043	
		(0.004)			(0.018)	
		[-0.014,0.002]			[-0.081,-0.012]	
$\ln H_0$ x $\ln \theta_{N,0}$			-0.010			-0.082
			(0.004)			(0.029)
			[-0.018,-0.003]			[-0.143,-0.032]
N-Children	8656	8656	8656	8656	8656	8656
N-Classroom	1118	1118	1118	1118	1118	1118

Source: ECLS-K:1998-99.

Notes: Classroom-clustered bootstrapped standard errors and 95% confidence intervals are in parentheses and brackets respectively. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table C-4: Descriptive Statistics at Kindergarten Entry (ECLS-K vs STAR)

Variable	ECLS-K	STAR
A: Characteristics of Child		
Number of children	8,656	6,325
Age	5.68	5.74
Fraction male	0.51	0.51
Fraction White	0.68	0.67
Fraction Black	0.14	0.33
Fraction Hispanic	0.05	0.00
Fraction other Race/Ethnicity	0.13	0.00
B: Characteristics of Classroom		
Number of classrooms	1118	325
Average Class size	20.30	20.34
Fraction Teachers Female	0.99	1.00
Average teacher's years of experience teaching K	9.44	9.26
Fraction of teachers having at least an MA	0.33	0.35
C: Characteristics of School		
Number of schools	637	79
Fraction public school	0.66	1.00
Fraction rural school	0.25	0.46
Fraction inner-city school	0.38	0.23
Fraction urban/suburban school	0.37	0.31
Fraction of students receiving free or reduced price lunch	0.25	0.48

Source: ECLS-K:1998-99 and STAR.
Notes: This table shows the comparison in the sample characteristics between the ECLS-K and STAR data.

Table C-5: Estimate for VA Models

	Math	Reading
Math (t-1)	0.653	0.237
	(0.011)	(0.010)
Reading (t-1)	0.160	0.517
	(0.011)	(0.010)
Child's Race: Black	-0.235	-0.070
	(0.036)	(0.034)
Child's Race: Hispanic	-0.113	-0.003
	(0.038)	(0.035)
Child's Race: Hispanic (Race not Identified)	-0.079	-0.044
	(0.046)	(0.043)
Child's Race: Asian	0.009	0.155
	(0.045)	(0.042)
Child's Race: Native Hawaiian	-0.062	-0.026
	(0.108)	(0.101)
Child's Race: Native American	-0.104	-0.095
	(0.092)	(0.085)
Child's Race: More than One	-0.118	0.066
	(0.050)	(0.047)
Free Lunch	-0.120	-0.141
	(0.031)	(0.029)
Reduced-Price Lunch	-0.063	-0.125
	(0.042)	(0.039)
SD Value-Added Estimates	0.323	0.395
Observations	8656	8656

Source: Model estimates using a sample of ECLS-K data.

Notes: Classroom-clustered bootstrapped standard errors are in parentheses. All models control for gender, age, age squared, and the time difference between the Fall and Spring assessments.

Table C-6: Analysis of Classroom Quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Math Classroom Quality			Reading Classroom Quality			Noncognitive Classroom Quality		
Number of hours/day	0.099	0.057	0.053	0.153	0.222	0.221	-0.030	0.051	0.052
	(0.021)	(0.095)	(0.095)	(0.020)	(0.065)	(0.065)	(0.021)	(0.077)	(0.076)
Class Size	-0.007	0.015	0.015	-0.011	-0.047	-0.047	0.011	0.012	0.012
	(0.008)	(0.026)	(0.026)	(0.008)	(0.021)	(0.021)	(0.008)	(0.028)	(0.027)
Teacher's Age	0.005	-0.000	-0.000	0.010	0.007	0.007	0.003	0.003	0.003
	(0.004)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)
Years of Experience at the current school	-0.015	-0.015	-0.015	-0.010	-0.004	-0.004	0.002	-0.009	-0.009
	(0.005)	(0.009)	(0.009)	(0.005)	(0.007)	(0.007)	(0.005)	(0.010)	(0.010)
$\ln H_0$			-0.015			-0.011			-0.012
			(0.004)			(0.003)			(0.004)
$\ln \theta_{M,0}$			0.000			0.000			-0.011
			(0.009)			(0.007)			(0.011)
$\ln \theta_{R,0}$			0.030			0.007			-0.002
			(0.010)			(0.008)			(0.011)
$\ln \theta_{N,0}$			-0.023			0.006			-0.002
			(0.009)			(0.008)			(0.010)
N	7552	7552	7552	7552	7552	7552	7552	7552	7552
School F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Source: the ECLS-K and STAR data.
Notes: this table shows how the estimated latent classroom quality is associated with teachers' and children's characteristics.