Racial Gaps in Cognitive and Noncognitive Skills: The Asian Exception

Christian Goldammer
University of Chicago*

July 20, 2012

Abstract

In recent cohorts, Asians have better economic outcomes than whites. Can this advantage be explained by early-life skills? This paper documents that cognitive skills are likely *not* the explanation: By 5th grade, Asians' cognitive skills (measured via test scores) are similar to whites'. In contrast, blacks and Hispanics have strong gaps in cognitive skills. This puzzle can partially be resolved by allowing for noncognitive skills (as rated by teachers): Asians exceed whites by .5 sd. Using an external dataset, this paper predicts that Asians' high noncognitive skills will translate into a 2.5 log point advantage in wages.

To understand the development of the skill gaps between Kindergarten and 5th grade, this paper estimates a dynamic skill formation model. For blacks and Hispanics, skill gaps are attenuated after controlling for inputs. In contrast, the noncognitive advantage of Asians *increases* after controlling for inputs.

*cgoldammer@uchicago.edu. I am indebted to James Heckman, Steven Durlauf and Matt Notowidigdo for their generous advice. In addition, I have received valuable comments from Derek Neal, Philipp Eisenhauer, Snaebjorn Gunnsteinsson, Tim Kautz, Miriam Gensowski, and many others. Furthermore, the participants of the CEDEC workshop have given me tremendously useful input.

1 Introduction

In recent cohorts, US-born Asians exceed whites in economic performance. Given this advantage, it is natural to ask how such a gap can be explained. This paper seeks to shed light on this fact by analyzing early-life skills using the ECLS-K.

Two pictures tell a thousand words, and they are given in figure 1, which presents raw cognitive and noncognitive skill gaps for blacks, Hispanics and Asians. These gaps are expressed in terms of the underlying skill factors. In cognitive skills, Blacks start from a low baseline, and then drop further, ending up with a gap of more than 1 sd. Hispanics make up for some of their initial gap, and Asians have gaps close to zero by fifth grade.[1] For noncognitive skills, however, the situation looks very different: The gaps of Hispanics are close to zero. The gaps of blacks increase over time, but recover by 5th grade. Asians, on the other hand, show continually rising performance.

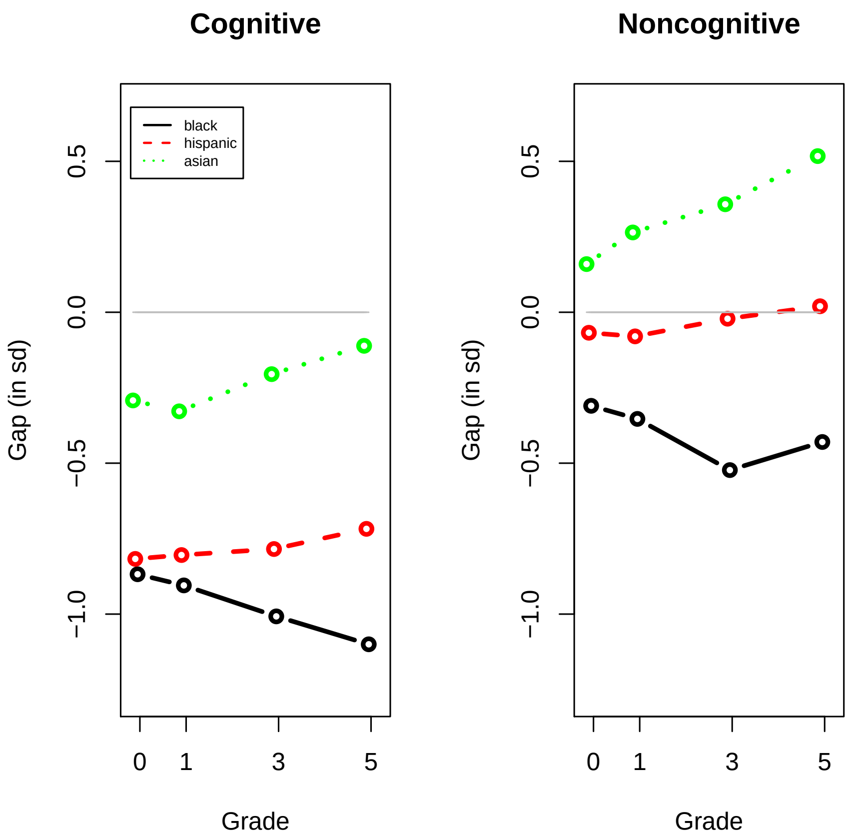

Figure 1: Skill Gaps, raw

This paper starts with a brief description of racial gaps in economic outcomes using the ACS. In recent cohorts, US-born Asians exceed whites in education by more than one year, and their hourly wages are higher by 20% or more. Wage gaps for Asians are strongly affected by controls for location, but gaps are positive in all specifications. In addition, the paper asks whether crime outcomes differ by race. Recent data shows that Asians are less than half as likely as whites to get arrested.

This paper continues by moving to the ECLS-K and displaying the racial gaps in early-life skills

[1]They lag by approximately .3 sd in Kindergarten, but this number depends, as explained later, on the imputation scheme for missing values.

and investments. Descriptive statistics highlight that, for noncognitive skills, it matters whether one asks parents, teacher, or the children themselves: Racial gaps are inconclusive if measures by parents or children are used. However, a clear ranking emerges using teacher ratings. In the causal analysis, I use teacher ratings, because these are available longitudinally and are of high data quality.

To answer whether differential paths can be explained by inputs, this paper applies the skill formation model of Cunha and Heckman (2008). Specifically, I calculate the factor loadings and the correlations of skill factors. All models allow for home investment and parental background. Noncognitive skills are more strongly affected by inputs than cognitive skills. School quality has a small and inconsistent causal impact.

With the production function identified, it is possible to calculate counterfactual racial gaps. One counterfactual starts from the initial race gap in Kindergarten: Given this gap, and the inputs along the way, are the observed racial gaps higher or lower than could be expected? Such a counterfactual equalizes the baseline performance and identifies whether races differ in their trajectories.[2] These gaps are presented in figure 12, which isolates the outstanding performance by Asians on noncognitive skills.

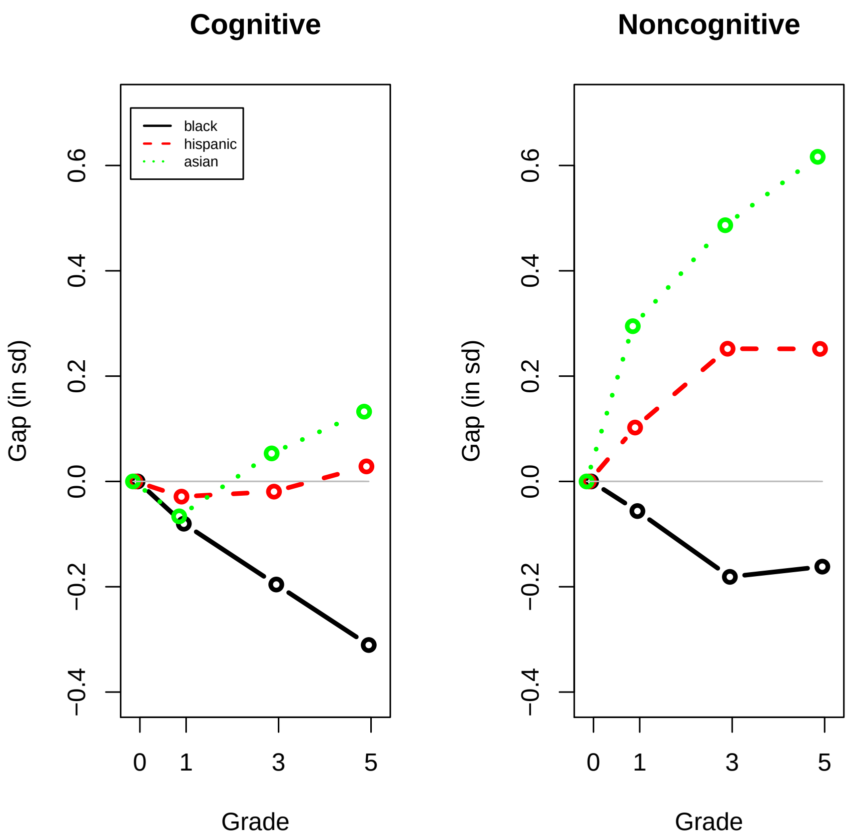

Figure 2: Skill Gaps, equalizing inputs and initial skills

Finally, this paper seeks to estimate whether skill gaps can explain outcome gaps. Doing so requires a dataset that contains information on childhood skills and later-life outcomes. I perform this analysis using British children born in 1970 (the BCS70 data).[3] For Asians, I find that their noncognitive skill advantage translates into a wage advantage of 2.5 log points.

[2]In no analysis do I intend to explain the initial skill gaps via controls. This is problematic because early gaps are produced from early-childhood experiences and inputs, of which the survey contains only limited information.

[3]Since the estimation sample (British children born in 1970) and the forecasting sample (Children born in the US in the

The paper is organized as follows. Section 2 gives a brief overview of the literature on skill gaps. Section 3 estimates racial gaps in education and wages using the ACS. Section 4 describes the main data, the ECLS-K. Section 5 presents racial gaps in skills and investment measures. Section 6 describes the factor model. Section 7 presents descriptive evidence on the factors. Section 8 estimates a dynamic model of skills. Section 9 calculates (raw and counterfactual) racial skill gaps. Section 10 estimates the impact of early-life skills on economic outcomes using the BCS70. Section 11 concludes. The appendix covers technical aspects.

2 Background

Skills and Gaps Economics has long recognized that there are substantial racial differences in economic outcomes (Altonji and Blank, 1999). It is also known that, conditioning on background characteristics, racial differences in cognitive tests by youth can explain a large part of the black-white gap in wages (for instance, see Neal and Johnson, 1995, who explain the wage gap for blacks using juvenile AFQT scores). Therefore, understanding the formation of skill differences is crucial to understanding racial differences in outcomes.

Here, I put emphasis on noncognitive skills. Recent research has shown that many of these skills and traits, measured in youth, are predictive of later outcomes (Almlund, Duckworth, Heckman, and Kautz, 2011). There exists little research on racial differences in noncognitive skills. When the measures for noncognitive skills are parental behavior ratings from the CNLSY, racial gaps are small for blacks and Hispanics after background characteristics are controlled for (car). Here, I document that racial gaps in noncognitive skills depend on the measures used: Using teacher ratings, gaps are large; there is no pattern when using parent or child ratings.

There exists recent work that complements my results. Using the ECLS-K, Claessens, Duncan, and Engel (2009) ask how much of later child performance can be explained via Kindergarten skills. Another recent paper uses a methodology similar to mine (DiPrete and Jennings, 2011) to investigate the gender gap in behaviors. Alongside other findings, their paper notes that Asians are rated highly by teachers on some teacher ratings.

To control for differences in inputs, this paper includes a skill model for skill formation. I apply the linear factor model introduced by Cunha and Heckman (2008) and introduce only slight modifications. Their application builds on a long literature that seeks to estimate the production function for skills. An overview over specifications and identification challenges is given by Todd and Wolpin (2003).

Outcomes of Asians It is well known that Asians achieved high levels of education (Sue and Okazaki, 1990). Recent evidence by Sakamoto, Goyette, and Kim (2009) shows that Asians also perform well on other economic outcomes, such as wages. Their paper documents that the economic performance differs widely by immigrant group: Chinese, Japanese, Indians, and Koreans have better outcome than whites. In contrast, a few groups (Hmong, Laotians and Vietnamese) lag strongly to whites. A likely reason is that a large part of the latter group immigrated as refugees, and thus were not selected on economic performance.

3 Racial Gaps in Outcomes

Here, I briefly document racial gaps in education, wages, and arrests (a measure of criminal behavior). This is meant to establish a basic fact: Among younger workers (below age 45), US-born Asians perform better than whites. For an overview over the interpretation of log wage gaps, see Altonji and Blank (1999). I use data from the ACS surveys between 2006 and 2010, and restrict the sample to individuals who are white, black, Hispanic or Asian and between the ages of 25 and 45. In wage regressions, I further restrict the sample to respondents that are not in school. When calculating skill gaps, I use a sample of

mid-90s) differ strongly, such a forecast needs to be treated with caution. However, the forecasting exercise is transparent and useful: It allows the reader to adjust the results based on his priors about wage impact of early-life skills.

students that are born in the US. For this reason, I further restrict the sample to US-born respondents. All results are weighted using the household weight (hhwt).

To document racial gaps in crime rates, I use recent data on arrests provided by the FBI.[4] The data provides raw arrest data pooling both genders and only provides two age groups: juveniles (below age 18) and adults (above age 18). Since the focus in this paper is on young individuals, I restrict the sample to juveniles. To convert total arrests into relative rates, I divide the arrests by the number of juveniles between age 10 and 18, assuming that children below 10 do not get arrested.

Education Figure 13 presents racial gaps in years of education for women and men, the numbers are given in table 24. Start with the first panel, for women. Blacks lag by $.74 - .95$ years, and Hispanics lag by $1.1 - 1.3$ years. Asians, on the other hand, exceed whites by at least 1 year. For Asian men, the gaps are even larger.

Wages Figure 14 presents racial gaps in log wages (numbers in table 25). These numbers are obtained from a regression where only time dummies are used as additional controls. The solid lines are raw gaps. I start with women: Blacks lag by between 14 and 20 log points, Hispanics lag by between 9 and 14 log points. Asians outperform Whites by between 25 and 34 log points for Women, and between 15 and 25 log points for men.

The dashed lines show how much of the gap is explained via a linear control for years of education. For blacks, this covariate explains roughly a quarter of the gaps, and it explains most of the gaps for Hispanics. For Asians, education explains roughly 40% of the gaps for women and more than half of the gap for men.

In unreported regressions, I find that the estimated gaps for Asians are strongly affected by controlling for geographical location, because Asians are likely to live in high-wage locations (such as California). For instance, including a control for region of residence reduces the gaps by approximately 30%. Including controls for state of residence area reduces the gaps by approximately 50%. However, even when this endogenous variable is controlled for, wage gaps still stand between 5 and 20 log points. If instead controlling for state of birth, wage gaps are only weakly lowered.

These numbers make a simple point: In recent cohorts, US-born Asians outperform whites in education and earnings.

Arrests Table 26 presents arrest rates among juveniles. Arrest rate are provided as arrests per 100,000 juveniles. The first row shows total arrests (sum of arrests over all categories). In this year, the arrest rates were 3.8% for whites, 8.4% for blacks, but only 1.4% for Asians. This implies that Asians were only 37% as likely as whites to get arrested. The remaining rows show arrest rates for specific crime categories. For readability, the list provides only the most salient crime type (the first 10 types in the full table), but the results look similar for the remaining types. For each crime type, Asians are less likely to be arrested. The relative rate ranges from .25 for burglary to .58 for robbery.[5]

Just as for economic outcomes, the results from the FBI data are clear: Asians are less likely to be arrested than whites. The data are far from perfect: Instead of using arrest rates as outcomes, it would be preferable to use convictions (because these are more likely to correspond to the individual and social cost of crime). For instance, it is possible that differential racial arrest rates are partially explained by differential police enforcement for the same behaviors. In addition, the data do not allow a restriction to just US-born Asians. Thus, the lower arrest rate for Asians can potentially be explained by the behavior of immigrants, who are afraid of being deported for committing crimes.

4 Data

[4] The data can be found at `http://www2.fbi.gov/ucr/cius2009/data/table_43.html`.

[5] For Asians, relative rates are relatively stable between crime types. For blacks, however, there are dramatic differences: Overall, they are 2.21 as likely as whites to get arrested. For robbery, however, they are 10 times as likely to get arrested.

Wave	Time	grade in school	number of observations	median age
1	Fall 1998	Kindergarten	21260	5
2	Spring 1999	Kindergarten	20649	6
4	Spring 2000	1	17487	7
5	Spring 2002	3	15305	9
6	Spring 2004	5	11820	11
7	Spring 2007	8	6878	14

Table 1: Sample size in the ECLS-K

Grade in school	Attrition in %	% of attrition attributable to subsampling movers
1	15	45
3	12	38
5	23	17
8	42	0

Table 2: Attrition in the ECLS-K

I use the Early Childhood Longitudinal Study Kindergarten Class of 1998-99 (ECLS-K). The ECLS-K was instituted by the Department of Education and extends from Kindergarten to 8th grade, starting out with 21,260 kindergartners and spanning an age range of 5-14 years. The data contains a large number of variables measuring child achievement, school quality, and home investment. Importantly, the ECLS-K oversamples Asians/Pacific-Islanders, which increases the numbers of Asian students that can be used for this analysis.

So far, the ECLS-K includes seven waves of data collection, as laid out in the table below. I omit the experimental wave 3 that uses only a subsample. In addition, I restrict the sample to US-born children. This eases the interpretation, because the circumstances of children immigrating during infancy potentially varies widely. However, I have run the analyses on the full sample, which does not affect the results.

The number of observations is given in table 1. Worryingly, attrition is high. However, not all of it is selective: a specific feature of the ECLS-K is that, in most waves, the students that switched schools were subsampled. This subsampling cannot lead to selection bias if it is corrected for by using sampling weights. As table 2 shows, this procedure explains a good share of the attrition, but less than half of it. Attrition is very high (42%) for the grade-8 sample, of which none can be explained by sub-sampling. Earlier analyses have found that grade-8 data quality is generally low. For this reason, this analysis does not consider grade-8 outcomes.

In the appendix, I further explain the sampling scheme and describe (non-conclusive) tests for selective attrition.

5 Racial Differences in Inputs and Skills

The ECLS-K contains an extensive set of skill and investment measures.

5.1 Skills

Later, when estimating causal effects, I will not strictly assign measures as cognitive or noncognitive, but let the factor model decide. However, to present skill gaps, I group measures as either cognitive or noncognitive. Table 5 presents the range of measures available.

5.1.1 Measures

Cognitive

- Test scores: Children in all waves were tested for math and reading skill. The tests were based on the NAEP Reading Frameworks, and allow separating key skills (Najarian, Pollack, and Sorongon). In addition, the survey measures a score for acquired knowledge (replaced by a test of science knowledge by grade 3 and 5). There were multiple skill levels, and test scores are provided in IRT (Item Response Theory) format, which attempts to compare responses across different assessment forms. Here, I use these IRT scores.
- Teacher rating of skills: Waves K-5 contain a teacher rating of child performance in both math and reading. In addition, they contain a teacher evaluation of the student's approach towards learning.

Noncognitive

- Teacher rating of personality: These measures include externalizing and internalizing behavior, self-control and interpersonal skills. All these measures are predicted factor scores from a factor analysis of a teacher survey on the children's social behavior. Because their definition is crucial for this paper, I cite the variable description from the manual (Tourangeau et al., b).
  - *Self-control* "has four items that indicate the childs 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."
  - *Interpersonal Skills* "has five items that rate the childs skill in forming and maintaining friendships ...".
  - *Externalizing Problem Behaviors* "include acting out behaviors. Five items on this scale rate the frequency with which a child argues, flights, gets angry, acts impulsively, and disturbs ongoing activities."
  - "The *Internalizing Problem Behavior Scale* (Teacher SRS) asks about the apparent presence of anxiety, loneliness, low self-esteem, and sadness. This scale comprises four items"
- Self-rating of the child: Externalizing, internalizing, and social behavior. These scales are developed similarly to the teacher ratings. They are available only from grade 3 onward.
- Parental rating of child: Self-control, social skills, sadness (introversion), impulsivenss (extraversion). These ratings are available for Kindergarten and grade 1. It has been noted that data quality is poor for these measures (DiPrete and Jennings, 2011).

5.1.2 Racial Gaps

The racial gaps are given in figures 3-6. All figures are given in the same format. There are three racial groups: black (bold line), Hispanic (dashed), Asian (dotted). All measures are normalized, thus the graph displays the gaps (in sd) of that racial group compared to whites.

Figure 3 shows gaps for measures of cognitive skills. In the test score measures for reading and math, blacks lag in Kindergarten, but then fall behind even further, ending at roughly one sd difference. Equally dramatic are the gaps when using the knowledge test. In contrast, gaps are smaller when using teacher ratings: The gaps from Kindergarten stay roughly constant for measures of literacy and math. Hispanics perform below whites, but do not show a dramatic drop in test scores. The measures for Asians are somewhat contradictory: While the gaps is close to zero for test scores, it continuously improves when students are rated by their teachers. This suggests that these measures represent different factors. The factor model will explicitly allow for this possibility.

Figure 4 shows gaps for teacher measures of noncognitive skills. Gaps are small for Hispanics, mostly large and negative for blacks and positive for Asians. Again, there is an interesting pattern: The skills of blacks decrease modestly over time, while Asians improve strongly.

Figure 5 shows gaps for parental ratings of noncognitive skills. Interestingly, these are strongly at odds with teacher ratings: For instance, parents of black children rate that their kids have a social skill gap close to zero by Kindergarten, while Asians are rated with a gaps of more than .5 sd.

Figure 6 presents the students' self-rating of noncognitive skills. There are no clear patterns: By third grade, Blacks have large gap in externalizing and internalizing behavior but rank higher than whites in peer behavior, while Asians have a gap only in peer behavior.

5.2 Investment

5.2.1 Measures

All measures are given in table 6. Due to the high number of survey questions, some of these are collapsed into indices (bold font). I use the following set:

1. The number of children's books in the household
2. An index of parental activities with children
3. Index of parental school participation
4. Index of child participation in school activities
5. Hours of TV watched by the child (inverted)
6. Number of library visits per month
7. Number of weekly meals that are held jointly

5.2.2 Gaps

The raw gaps are given in figure 7. The strongest gaps can be seen for the number of books, with blacks and Hispanic initially being more than 1 sd behind white; however, the gaps decreases to around $-.8$ over time. Asians start around .4 sd higher and move in a similar trajectory. In most measures, blacks invest less than whites. An exception are the activities with children: By third grade, blacks invest more than all other groups, including whites. Asians have small gaps for some measures (TV, meals together), invest more than whites in library visits, and lag strongly in school involvements and child activities.

5.3 Background

5.3.1 Measures

For the parental background factor, I use the following measures

1. Family income (log)
2. Socioeconomic status (generated by data providers)
3. Mother has white-collar job

5.3.2 Gaps

The raw gaps are given in figure 8. Blacks have strong gaps on all measures, especially on log income. Asians perform on par with whites, and Hispanics are similar to blacks, though with slightly higher income.

5.4 School quality

5.4.1 Measures

1. School Average: misbehavior (teacher-rated)
2. School Average: school spirit (teacher-rated)
3. School Average: students capable (teacher-rated)
4. School Average: Teacher's motivation
5. School crime problems index
6. Facilities index
7. Class size
8. Teacher's years of education
9. Teacher's work hours

The measures are given in table 7. The school averages are obtained by averaging all teacher measures from a school. These capture only a subset of the data given in the ELCS-K. As shown later, the explanatory power of the school quality factor is small. I have experimented with a wide range of additional measures, with equally dissatisfying results.

5.4.2 Gaps

Figure 9 gives the gaps. In contrast to the investment measures, these plots tell a clear story: Asians go to good schools (gaps are around zero or positive). In contrast, the schools of blacks lag in school problems (.5 sd by 5th grade), in the teacher motivation (.1 by 5th grade) and in class behavior (.45 by 5th grade). Blacks do not lag when it comes to small classes or the number of aides, but these measures are problematic: Both can reflect low demand for a school, and thus be perversely correlated with school quality. One interesting exception is that blacks have small gaps in school facilities, while gaps for Asians and Hispanics stand at around .7 sd by 5th grade.

6 Statistical Model

There exists an extensive literature that seeks to establish the production function for skills. A recent overview is provided by Todd and Wolpin (2003). The following is an application of the skill formation model of Cunha and Heckman (2008). It offers several features:

- Investments and skills are allowed to be measured with error. The factor structure allows correcting for the resulting bias. This does have considerable effect on parameter estimates: If inputs is correlated with skills, then measurement error in inputs will lead to an upward bias in the estimated strength of investment.
- Investments are allowed to be endogenous. Identification is obtained using (assumed to be exogenous) lagged investment.
- Skills are self-productive. Therefore, an investment in period t will affect skills in all future periods. The fade-out of investments is identical to the fade-out of skills. This contrasts with value-added specifications, in which only contemporaneous inputs are allowed to affect skills. However, it does restrict the fade-out process for investment.

- The number of causal factors is small. This stands in contrast to applications of OLS regressions, which often contain dozens of explanatory variables. It is, however, problematic to interpret all these variables causally. For instance, assume that the paper used a large number of covariates for investment, and instrumented each variable using its lagged value. This procedure does not yield consistent results in the presence of measurement error. To avoid this, this paper goes the alternative route, by collapsing a large number of measures into a small set of causal factors, which reduces the econometric problems in the resulting model.

- The model does not allow for underlying factors (such as genetic effects) that affects skills in every period. Instead, a central model assumption is that any fixed effect operates *only* through initial skills. For this reason, initial skills are treated as a black box.

In a recent paper, Todd and Wolpin (2007) compare different specifications. They find that cognitive skills are affected both by contemporaneous and lagged inputs, using a single measure of inputs. Their work contrasts with Fryer and Levitt (2006), who use only contemporaneous inputs as control variables. An earlier version of this paper used a large number of covariates, but this changes racial gaps little compared to a more parsimonious specification. However, allowing for measurement error does affect coefficients, because home investment is strongly correlated with skills. The factor model is a convenient method to address measurement error.

In addition, the factor model provides a consistent way to understand the large number of measures available in the data. The factor model asks for the structure underlying the measurements, and uses the underlying correlations to estimate causal effects. This prevents data-mining for the covariates that predict skills most strongly, which might leave the econometrician with a heap of variables that are strongly predictive but likely endogenous.

6.1 Basic Setup

There are T stages. In each stage, a student is described by J factors θ_t^j and stacking the factor yields $\theta_t \equiv \left[\theta_t^1, \ldots, \theta_t^J\right]'$. The factors evolve according to the equation

$$\theta_t = G_t \theta_{t-1} + \eta_t$$

with $V(\eta_t) \equiv Q_t$. The latent factor θ is multi-dimensional and can include not only skills, but different types of investment, and allow for their endogeneity. G_t is the main matrix of interest, it tells all causal stories of interest. If one defines $\gamma_t^{jj'} \equiv G_t\left[j, j'\right]$, the regression equation for an individual factor can be written as follows:

$$\theta_t^j = \sum_{j' \neq j} \theta_{t-1}^{j'} \gamma^{jj'} + \eta_t^j$$

I split the factors j into two groups: Skills and Investment. The measurement equation is different for these sets of factors.

For each investment factor j and period t, one observes m_t^j measurements for each factor, called $Y_{t,k}^j$ with $k \in \left\{1, \ldots, m_t^j\right\}$. The measurement is related to the factor via the equation

$$Y_{t,k}^j = a_{t,k}^j \theta_t^j + e_{t,k}^j$$

Where e is measurement error. Note that $a_{t,k}^j$ and θ_t^j are both scalar. The measures are stacked as

$$Y_t^j \equiv \left[Y_{t,1}^j, \ldots, Y_{t,m_t^j}^j\right]$$

For the skill factors, one observes m_t^S measures. Denote the vector of skill factors by θ_t^S, then the measurement equation is given by

$$Y_{t,k}^j = a_{t,k}^S \theta_t^S + e_{t,k}^j$$

where $a^S_{t,k}$ is now a vector.

All these measurement equations can be stacked by defining Y_t as the set of all measures in t:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

Where the α_t is restricted to allow associating measures with factors.

All measures are assumed to be normally distributed. This is not necessary for identification, as the distributions of the factor and the errors are nonparametrically identified Cunha and Heckman (2008). However, any relaxation is numerically intensive and prone to convergence problems. In fact, the model I estimate already hits the computational limits of my computer.

Regarding the measurement errors, I originally made the same assumptions as the original paper, thus estimation can be performed via the Kalman filter. I programmed this, but the procedure lead to problems: Specifically, I obtained that some diagonal elements of Q_t equal zero (that is, they are hitting the constraint that variances have to be non-negative). Therefore, I am not insured that this leads to consistent estimates of G_t. For this reason, I estimate a (more general) structural equation model that allows for arbitrary serial correlation between measures. When I present the results, I will explain the model in more detail.

6.2 Factor Structure

For the investment factors, measures are associated with a unique factor. Identification of causal parameters is assured whenever there are more then three measures, but factors require a normalization. In the measurement equation

$$Y^j_{t,k} = a^j_{t,k} \theta^j_t + e^j_{t,k}$$

I require that $a^j_{t,1} = 1$: The first loading is normalized to unity.

For the skill factors, I allow a measurement to be affected by multiple factors. To label factors, it is necessary to restrict the factor structure. Here, I assume that there exists at least one measure that loads on only a single factor. For instance, in the application, I will specify that a student's test score loads only on the cognitive factor. However, teacher evaluations of learning are allowed to load both on the cognitive and noncognitive factor. The use of dedicated measures provides a clear interpretation to the factors and eliminates the need for factor rotation that is common to factor analysis.

6.3 Extension: Contemporaneous and Endogenous Inputs

The statistical model used so far is given by

$$\theta_t = G_t \theta_{t-1} + \eta_t \tag{1}$$

I now extend this model and allow for some factors to affect other factors contemporaneously (and not only in the future). To see why this is plausible, assume that there are two factors: cognitive skills S and investment I, with I not being causally affected by skills. However, current skills are affected by current investment (and not lagged investment). In other words:

$$\begin{pmatrix} \theta^C_t \\ \theta^I_t \end{pmatrix} = \theta_t = \begin{pmatrix} \gamma^1_{11} & 0 \\ 0 & 0 \end{pmatrix} \theta_{t-1} + \begin{pmatrix} 0 & \gamma^2_{12} \\ 0 & 0 \end{pmatrix} \theta_t + \eta_t$$

so the more general notation is to write

$$\theta_t = G^1_t \theta_{t-1} + G^2_t \theta_t + \eta_t$$

where some caution is required: the model can be rewritten as

$$\begin{aligned} \theta_t &= \left(I - G^2_t\right)^{-1} \left(G^1_t \theta_{t-1} + \eta_t\right) \\ &= \widetilde{G}_t \theta_{t-1} + \widetilde{\eta}_t \end{aligned}$$

and the intuitive criterion for identification is that the number of estimated parameters in G_t^1 and G_t^2 is less than the number of elements in $\widetilde{G}$.

As before, the covariance structure of the shocks η_t is not restricted. This allows for inputs to be endogenous (because the shock to investment might be correlated with shocks to skills). In the next subsection, I provide an intuition for the identification within this model.

6.4 Identification Assumptions

Above, I claimed that this model allows making causal statements. Causal statements, however, are not true per se, but are a statement within a model. Here, I wish to add (mostly) intuition on the causal claims made by the factor model, and highlight how it compares to more standard panel data methods. I will assume two factors to simplify exposition: skill and investment. Let the true model be given by

$$\theta_t^S = \gamma_t^1 \theta_{t-1}^S + \gamma_t^2 \theta_t^I + \eta_t^S \tag{2}$$

Intuition on Identification of Parameters The intuition can be best conveyed by assuming that, for each factor and time period, there exists a measure that is perfectly correlated with the factor. In that case, the factors are (up to normalization) perfectly observed. Using the normalization, it is possible to directly estimate the production function for skills (2). Does OLS reveal the parameters γ? It does so only if

$$Cov\left(\theta_t^I, \eta_t^S\right) = 0$$

However, the dynamic model does not requires this, because it does not restrict the variance matrix of η. In other words, the model allows for endogeneity of the investments. How then is the causal parameter γ^2 estimated? Intuitively, it is identified because past investment (which is uncorrelated with η_t^S) serves as an instrument for current investments. This is a common identifying assumption for panel data methods. Clearly, it is not without problems: it assumes that households do not act based on information about future shocks that are unobserved to the econometrician.

Full Specification Required Another potential problem is that the model estimates all causal parameters jointly. This requires specifying all aspects of the model. For instance, if the econometrician is interested in the parameter γ_t^2 (of investment on skills), it is not enough to just specify equation (2). Since investment is a factor, the econometrician has to also specify how investment depends on (either past or contemporaneous) skills. A misspecification in one equation will lead to inconsistencies in all other equations.

While this is a theoretical concern, in practice this does not seem to matter much in my application: for instance, I have estimated the self-productivity of cognitive skill using a wide range of assumptions about the causal effect of skills on investments, and the estimates are rarely affected by different specifications.

Joint Normality A functional form for the error terms and factors is not required for identification, but commonly invoked to ease estimation. In this application, the model is estimated assuming joint normality of all factors and measures.

7 Understanding the Factors

A *factor* is an underlying trait that is unobserved, but is imperfectly measured. In this section, I present the loadings for the factors used in the analysis:

- Skills
- Parental Investment

- Parental background
- School Quality

7.1 Skills

Factor Loadings As laid out in the last section, I allow for measures to be affected by both cognitive and noncognitive skills, but I ease interpretation by anchoring cognitive skills in the test scores.

I start by using a model that includes all measures that are available for certain waves. The waves K-1 contain test scores, measures from teachers and from parents. I restrict the number of factors to three, which is highly illustrative. I anchor the first factor in test scores, the second factor in the teacher measure of externalizing behavior and self-control, and the third factor in parental measures of internalizing. Table 8 presents the results. These tables are all in the same layout. They present the results for one or more factors, and all possible waves. Each cell contains the normalized factor loading (between -1 and 1) and the standard error in parentheses.[6] All measures are defined such that the factor loading is expected to be positive, thus negative factor loadings signal problems.

The cognitive measures are all highly correlated, factor loadings are above .8. Note that these measures are excluded from the noncognitive factors, thus interpretation is straightforward. Teacher ratings split onto the first and second factor: Ratings of school performance (in math and reading) load onto the first factor, while rating of behavior load onto the second factor. Furthermore, parental measures load highly on the third factor, but not other factors. Thus I call the first factor "cognitive skills", the second factor "noncognitive skills, teacher-rated" and the third "noncognitive skills, parent-rated"

These results are not consistent with a model that posits that noncognitive skill is scalar and enters both parental and teacher ratings. It is also not consistent with separate noncognitive traits (such as externalizing and internalizing) that are observed by both parents and teachers. If this were true, measures of externalizing behavior would all load on a single factor.

The pattern is similar when including self-ratings by the child in waves 3-5, table 9. The noncognitive factors can be labeled as teacher ratings and self ratings.

For this reason, I do not proceed by including all measures. Because the teacher measures are consistent over the first four waves, I spend most time discussing the results using just teacher measures of noncognitive skills. However, I briefly describe the results when using parental or self ratings. The factor loadings when using just teacher measures are given in table 10. Overall, these are quite similar throughout the waves. The teacher skill ratings load more strongly on the cognitive than the noncognitive factor. The teacher evaluation of the student's learning skills loads more strongly on noncognitive skills. Among the teacher ratings, externalizing and self-control load highly on noncognitive skills, while the loading on internalizing behavior is lower.

Correlations of Skill Factors The factor analysis showed that noncognitive measures given by parents, teachers, and students originate in different factors. However, it is still possible that these factors are highly correlated. To shed more light on this, table 11 presents the intercorrelations of factors. The correlations between cognitive skills and teacher-rated noncognitive skills are low, all stand below .25. The correlations are higher on average (but more variable) between teacher and parent-rated noncognitive skills. This holds similarly for correlations between teacher and child ratings. Thus, there is clear evidence that noncognitive and cognitive skills can be distinguished, but it is not plausible to combine parental and teacher ratings.

These low (at least, lower than expected by the author) correlations between noncognitive factors are potentially the result of differential comparison groups used by teachers and parents. The teacher and parent measures are themselves predicted factor scores as generated by the data provider. The documentation does not provide the exact survey questions (see Tourangeau et al. (a)). Crucially, it is possible that teachers interpreted their answer relative to a baseline, for instance their class or their school. In an earlier version of this paper, I allowed teacher measures to depend on class behavior, but

[6]If a standard error is zero, this comes about because the specific parameter was normalized.

this model was hard to estimate and highly unstable due to high correlations of measures and class averages when only few students are sampled in each class.

How to interpret the differences in factor loadings and relatively low correlations? One approach would be to assume that all measures of noncognitive skill are produced by one underlying factor, and that different sources (teacher, parent, student) have measurement errors that are autocorrelated. This assumption, however, clashes with the finding that certain measures (social skills) load more heavily for teachers than for parents (not reported here, but in an earlier version of this paper).

An alternative explanation is that parent and teacher ratings of noncognitive skills differ systematically. It is possible that they observe different behaviors: A child plays a role at home, and a different role in the school. Noncognitive skills are thus partly a choice of the child, and can be expected to depend on its budget constraint and prices. The implications of this model are anything but obvious, but I will not analyze this possibility further in this paper.

7.2 Investment and Parental Background

The factor loadings for parental investment are given in table 12. All loadings are positive. Overall, the measures that load most strong on the factor are the number of books, parental school attendance, and parental participation. The loadings are relatively stable throughout the waves.

Table 13 presents the loadings for the parental background factor. These are all above .5, and highest for SES, at above .95. Since the factor and the measures are so strongly correlated, it is allowable to replace the factor by a principal components of its measures. I use this simplification of the model to ease the computational burden, but this does not noticeably affect the results in the specifications that I checked.

7.3 School Quality

The normalized factor loadings are given in table 14. There is little hint of one underlying factor. Most problematically, class size enters strongly in the unexpected direction. To allow interpretation, I reduce the number of measures to variables that are constant within the school (and thus can be interpreted as quality of the school). I present the results in table 15. Here, the loadings are positive (but close to zero for facilities). School quality loads most strongly on the absence of school crime problems.

8 Skill Formation

This section presents results from estimating the full causal model. The variables used change slightly from the pure factor analysis presented above. I remove the teacher ratings of skills (math and reading) from the analysis, because these skills were only measured in a subsample of the data. Including these measures reduces sample-size too much to obtain precise estimates of causal effects. For parental investment, I remove the measures of parental activities, library visits, TV hours, and meals together. Their factor loading are low (compared to the other measures), and the number of missing observations is high.

8.1 Models Without Investment

I start by estimating a baseline model that uses only two factors: cognitive and noncognitive skills (teacher-rated). Therefore, I avoid the problem of modeling multiple noncognitive skills, at the cost of sacrificing potential explanatory power.

The results are presented in table 16. This table presents the causal equation for skills, with outcomes in specific grades as the columns, and rows denoting explanatory variables. When presenting my results, I always normalize factor variances to unity to ease interpretation. In the following, I only present the matrix $\widetilde{G}_t$ where each coefficient gives the partial correlation between variables. The individual element is defined as

$$\widetilde{G}_t\,[j,j'] = \gamma_t^{jj'} \cdot \sqrt{\frac{V(\theta_{t-1}^{j'})}{V(\theta_t^j)}}$$

For instance, the coefficient of .97 of lagged cognitive skills on cognitive skills implies that a one sd increase in cognitive skills will increase future cognitive skills by .97 sd. For grade 3 and 5, the cross-effect from cognitive to non-cognitive skills is stronger than the other way, which stands in contrast to the finding of Cunha and Heckman (2008). An interesting aspect is that noncognitive skills are much less self-productive than cognitive skills: the autocorrelation stands at around .6.

This is a plausible baseline model, but the specification is too simple. An important statistical assumption is that measurement errors are uncorrelated over time. This is refuted by the data: Math scores are much more highly correlated with past math scores (.71) than with past reading scores (.52), with the reverse holding for reading scores. Similarly, the number of books (a component of investment), is very highly autocorrelated (above .8), but the correlation with other past measures is much lower (generally below .2). This assumption matters for estimating the parameters in G. Not allowing for this autocorrelation leads to an upward bias in self-productivity. The intuition is that, in the model (1), a high autocorrelation in measures can only be explained through high self-productivity of the factors. Autocorrelated measures provide an alternative explanation.

Thus I extend the measurement equations. For measure j, write the measurement error as

$$\varepsilon_t^j = \rho^j \varepsilon_t^j + \widetilde{\varepsilon}_t^j$$

In table 17, I allow for autocorrelated measurement errors, which affects the results: Self-productivity drops, both for cognitive and noncognitive skills.

8.2 Parental Investment

I now include factors for parental investment and parental background. In this specification, I assume that these factors are exogenous (which I later relax). Formally, I assume that the factor shocks η_t^I and the shocks to skills η_t^S are mutually independent.

The results are given in table 18. The causal effect for parental investment on cognitive skills is small, below .05 for all waves. However, the effects are much larger for noncognitive skills (around .1 in grades 1 and 3). It is important to note that investment is highly autocorrelated over time (above .9,not reported). Therefore, an early-life shock to investment has the potential to "stick" with a family and produce cumulative effects. Parental background has a more consistent effect: It affects both cognitive and noncognitive skills, with an effect size of approximately .03.

In the next specification, I allow investments to be endogenous by allowing free correlation between η_t^I and η_t^I. The results are given in table 19. Overall, the results look similar to table 18, which suggests that both parental investment and family background are not strongly endogenous — under the assumptions of the model.

8.3 School Quality

Finally, I include a factor for school quality in table 20. The effect sizes are widely varying for cognitive skills (positive in first grade, negative in third grade) and weak for noncognitive skills. I have used a wide range of specifications and measures for school quality. In cases were school quality predicts outcomes, the effect depends strongly on the specification used, and is negative as often as positive. This corroborates the work of Hanushek (1997), who finds that observable measures of school quality do not predict gains in test scores.

8.4 Other Noncognitive Factors

I now perform the same analysis for the noncognitive measures provided by parents (table 21). The cross-productivity is weak. Investment enters positively, background and school quality are not significant.

9 Counterfactual Skill Gaps

The last section estimated production functions for skill acquisition. Now, I return to racial skill gaps, and ask whether this model can help explaining them. Since the production function is identified, it is possible to construct counterfactual skill gaps. Here, I focus on skill gaps that equalize investments between racial groups.

9.1 Definitions

Skills are given by

$$\theta_t = G_t \theta_{t-1} + \eta_t \tag{3}$$

The RHS of this model contains not only covariates, but also past skills. Write $\theta = \left(\theta^S, \theta^I\right)'$. One can similarly split up the matrix G as

$$G_t = \begin{pmatrix} G_t^{S,S} & G_t^{S,I} \\ G_t^{I,S} & G_t^{I,I} \end{pmatrix}$$

In this model, a counterfactual skill gap is achieved by predicting from equation (3), removing the effects of the inputs (which differ by group). First, define total investment for race r as

$$I_t^{Total}(r) \equiv \sum_{\tau=0}^{t-1} \left[\prod_{\mu=1}^{\tau} G_{1,t-\mu}^{S,S} G_{2,t-\tau}^{S,I} E(\theta_{t-\tau}^I | r) \right]$$

The counterfactual expectation assumes that different racial groups have the same investment stream. For $t \geq 2$, this is:

$$E_t^{Conditional}(r) = E(\theta_t | R = r) - I_t^{Total}(r) \tag{4}$$

To see where this term comes from, note that skills at time t can be expressed as a function of initial skills θ_0^S and the whole stream of investments θ^I:

$$\begin{aligned}
\theta_t^S &= G_{1,t}^{S,S} \theta_{t-1}^S + G_{2,t}^{S,I} \theta_t^I + \eta_t &(5)\\
&= G_{1,t}^{S,S} \left[G_{1,t-1}^{S,S} \theta_{t-2}^S + G_{2,t-1}^{S,I} \theta_{t-1}^I + \eta_{t-1} \right] + G_{2,t}^{S,I} \theta_t^I + \eta_t &(6)\\
&= G_{1,t}^{S,S} G_{1,t-1}^{S,S} \theta_{t-1}^S + G_{1,t}^{S,S} G_{2,t-1}^{S,I} \theta_{t-1}^I + G_{2,t}^{S,I} \theta_t^I + \tilde{\eta}_t &(7)\\
&= \dots &(8)\\
&= \sum_{\tau=0}^{t-1} \left[\prod_{\mu=1}^{\tau} G_{1,t-\mu}^{S,S} G_{2,t-\tau}^{S,I} \theta_{t-\tau}^I \right] + \prod_{\mu=1}^{t} G_{1,t-\mu}^{S,S} \theta_0^S &(9)\\
&= I_t^{Total}(\theta^I) + E_t(\theta_0^S) &(10)\\
I_t^{Total}(\theta^I) &= G_{1,t}^{S,S} \left[I_{t-1}^{Total}(\theta^I) \right] + G_{2,t}^{S,I} \theta_t^I &(11)
\end{aligned}$$

The racial expectation can then be written as

$$\begin{aligned}
E^{Model}(\theta_t^S | r) &= I_t^{Total}(E(\theta^I | r)) + E(E_t(\theta_0^S | r)) &(12)\\
&\equiv I_t^{Total}(r) + E_0(r) &(13)
\end{aligned}$$

Please note: In contrast to other recent work (Fryer and Levitt, 2006), the initial skill gaps are not conditioned on. Therefore, my model takes the initial skill gap as a primitive, and can provide counterfactuals given this primitive.

Initial skills are allowed to be multi-dimensional. Theoretically, this can be important: If both noncognitive and cognitive skills affect each other, restricting the analysis to a simple factor could wrongly ascribe skill gaps. For instance, if blacks start with lower noncognitive skills than whites, and noncognitive skills positively affect cognitive skills, any regression that does not incorporate noncognitive skills will ascribe some of its effect to race. In my estimates, however, this is not likely to be a large problem: Cross-effects are too small to strongly affect racial gaps.

In addition to the counterfactual skill gap after equalizing investment, another useful counterfactual is the "Deviation from expectations". To understand this concept, consider the gap in cognitive skills for blacks by Kindergarten. Given the skill formation model, the initial gaps should diminish over time. Thus one can ask: How much larger are the skill gaps than they would be expected, given initial gaps and investments? This concept is given by

$$E_t^{Deviation}(r) \quad \equiv \quad E(\theta_t|R=r) - E^{Model}(\theta_t^S|r)$$

This is an overall metric of racial gaps, and I will use this to aggregate the model into a simple set of outcome statistics. It is useful for comparing time trends, since it incorporates the idea that initial gaps should revert to zero in the long run. Since all gaps start at zero, the Deviation counterfactual is a natural metric for divergence observed during the sampling period.

I estimate the factor model including cognitive and noncognitive skills, parental investment and background. I do not control for the noisily estimated school quality, but including it does not change the results.

9.2 Estimating Raw Racial Gaps

Given estimates of racial gaps in skills and investment, the causal model allows calculating counterfactual skill gaps.[7] It is not advisable to calculate raw gaps from the factor model that estimates the production function.

The first reason is that the factor model does not allow for sampling weights. Weighting the data is not trivial: The data providers produce weights that differ by outcome variable. For instance, the weight C1_6FC0 is appropriate for analyses that use respondents with observable child-level data from Kindergarten to fifth grade. The variable C1_6PC0 is appropriate when parent-level data is observed. However, there exists no weight to be used when both parental and child data are observed. Because my focus lies in child-level outcomes, I use the child-level longitudinal weights. Results are similar when using parent-level weights.

In addition, the sampling design leads to biases when estimating racial gaps within the model. This happens because children with low language skills might not complete a test score. Hispanics have high non-response to the reading and knowledge test in Kindergarten and first grade. This happens by design: All non-English speaking children completed an English screener. If scores were too low, the Spanish-speaking children received a translated version of the math test, but no reading and knowledge test. Asians who failed the screening have no test score available in Kindergarten. This biases race gaps estimated using the longitudinal sample: Children with low English skill (Hispanics and Asians) will be dropped.

I present the problem in table 3. The first two columns show the numbers of observations with and without reading test score in Kindergarten. For blacks the share without test score is small (below 1 percent), but not for Hispanics (.2) and Asians (.15). The next three rows look at the situation by first grade. Almost all these Asians *do* provide non-missing test scores by first grade, but there are still 10% of Hispanics who do not. What is the skill of those with missing information on test scores? These Hispanics and Asians lag by 1.4 sd! Thus, omitting these observations will produce a bias in estimating racial gaps. The final row provides a rough estimate of the bias: If the missing children were from a group with no racial gaps, how much would including them lower the gap? The number is between .2 and .3 for both Hispanics and Asians. This table makes one simple point: For cognitive skills, the racial gaps of Asians and Hispanics in Kindergarten will be biased if Kindergarten non-response is neglected.

I use a simple imputation procedure to alleviate this problem. It is important to note that the imputation is fully compatible with the factor model. Among Hispanics, those who fail the language

[7] This requires the assumption that the missing observations do not affect the estimate of G.

	white	black	hispanic	asian
Test Score Present, Kindergarten	19	15	519	118
Test Score Not Present, Kindergarten	9067	2097	1960	660
Share Missing Kindergarten	0.002	0.007	0.209	0.152
Test Score Present, Grade 1	1	3	251	22
Test Score Not Present, Grade 1	8914	2058	2177	732
Share Missing Grade 1	0	0.001	0.103	0.029
Cog. Skill (missing K)	-1.028	-2.626	-1.428	-1.408
Implied Bias (Cog)	-0.002	-0.019	-0.299	-0.214
NC Skill (missing K)	0.041	-0.3	-0.075	0.318
Implied Bias (NC)	0	-0.002	-0.016	0.048

Table 3: Gaps From Missing Test Scores in Kindergarten

test provide a valid math test. To obtain a score for their cognitive factor, I use the insight of the factor analysis. Let the math test score be given by

$$M^{math} = \alpha^{math}\theta^{cog} + u^{math}$$

Then, for racial group r, the racial gap in the cognitive factor can be calculated simply from the racial gaps in math scores, because

$$E(\theta^{cog}|r) = \frac{E(M^{math}|r)}{\alpha^{math}}$$

For Asians, this imputation is not possible. For those failing the language screener, all test scores are missing in Kindergarten. However, I note that their test scores are available by first grade. Thus I impute the Kindergarten cognitive score using their first-grade score. Again, look at the factor model. For simplicity, assume that there is no noncognitive factor. Then

$$\theta_t^{cog} = \gamma_{cog,cog}\theta^{cog} + \eta_t^{cog}$$

One can express Kindergarten scores using first-grade scores.

$$E(\theta_{t-1}^{cog}|r) = \frac{E(\theta_t^{cog}|r)}{\gamma_{cog,cog}}$$

It is not assured that the imputation scheme correctly captures the skills of imputed students. In that case, estimates of gaps in Kindergarten can be biased. Importantly, imputation does *not at all* affect the observed test scores by 5th grade. However, it does affect the sample, by allowing to include students whose Kindergarten scores were missing. Therefore, the imputation scheme might produce problematic numbers in Kindergarten, but it improves the estimated skill gaps in later grades.

Given the numbers in table 3, imputation should affect estimates of racial gaps. In table 4, I show that this is indeed the case. For Kindergarten cognitive skills, imputation increases the gaps of Hispanics by .13 and of Asians by approximately .18 sd. In 5th grade, the additional gaps are .16 for Hispanics and .11 for Asians.

	black	hispanic	asian
Kindergarten	-0.869	-0.688	-0.117
Kindergarten Imputed	-0.868	-0.817	-0.292
5th grade	-1.101	-0.541	0.007
5th grade Imputed	-1.1	-0.718	-0.111

Table 4: Impact of Imputation on raw gaps in Kindergarten cognitive skills

To estimate raw gaps, I use a sample which restricts all test scores to be non-missing. This is consistent with the used sample weight, which is based on longitudinal non-missing child observations. The total sample contains 8724 observations, out of which 465 are Asian. The sample size is lower when calculating teacher ratings, since these also contain missing values.[8]

For calculating racial gaps, it is important to distinguish between measures and factors. For instance, if blacks lag by 1 sd on all measures of cognitive skill, they will lag by more than one sd in the cognitive skill factor. This happens because there is noise present in the measures. To predict a factor score, I start by averaging all measures that load on a factor.[9] Let measures be given by $Y_1, \ldots, Y_K$, and factor loadings by $a_1, \ldots, a_K$. Then the normalized average of all normalized measures ($\bar{Y}$) can be written as

$$\bar{Y} = \bar{a}\theta + \tilde{e} \tag{14}$$

$$\bar{a} = \frac{\sum_k a_k}{\sqrt{\left(\sum_k a_k\right)^2 + \sum_i \left(1 - a_k^2\right)}} \tag{15}$$

The factor score is then defined as

$$\hat{\theta} \equiv \frac{\bar{Y}}{\bar{a}}$$

Mean differences in the racial score are consistent estimates of racial differences in standard deviations of the underlying factor.

To see an example: Assume that $K = 2$ and that $a = [.8, .8]$. Then the mean of all z-scores will have a SNR of .88. Thus, the factor score will equal the mean of all measures, normalized, and then divided by .88.[10]

9.3 Cognitive Skills

I present the results from all these models graphically in figure 10 (the precise numbers are provided in table 22). Standard errors are obtained via bootstrap with $R = 10$ iterations. For each race, the table presents raw gaps (specification 1), gaps when controlling for inputs (specification 2), and the deviation from expectation (specification 3).

Asians start with a lag of almost .3 sd. However, this number must be treated with caution, because it is heavily affected by the imputation scheme. By first grade, they fall behind somewhat more, but they recover by 5th grade, remaining with a gap of roughly .1 sd. Specification (2) allows for investments, which changes the results only little. Specification (3) shows whether Asians deviate from the expectation, given Kindergarten scores. They outperform their expectations by a little over .1 sd.

The skill gaps for black are much more dramatic. They start out lagging by .9 sd in Kindergarten and fall behind another .25 sd by grade 5. The second panel equalizes the investments between groups, which lowers the trajectory by roughly .1 sd. The final panel displays the deviance from expectations:

[8]Missing data is prevalent: The total number of observations is 19,657. The cognitive score for Kindergarten is available for 14,649 cases, and in grade 5 for 9,177 cases. As laid out above, part of this is explained via random sub-sampling, which can be corrected by using sample weights. After restricting the sample to have non-missing information in all waves and all factors, the sample size is reduced to 4153, with 165 Asians. Using this sample does not strongly change the estimated gaps in noncognitive skills, but it increases standard errors considerably. However, it considerably affects gaps in cognitive skills: This happens because a large number of students with imputed cognitive skills also have missing information on investment and noncognitive skills. However, gaps for investment and background are little affected. For this reason, using the sample with just cognitive skills is preferable to using a sample without any missing data. A useful consistency check for the results is to check whether the gaps are similar when using the cross-sectional data with cross-sectional weights. The variable C1PTW0 is appropriate for analyses of Kindergarten outcomes given that data on test scores, the parental survey, and the teacher surveys are non-missing. Since the skill factor requires inputs from both teachers and test scores, I weight them using C1PTW0. Another weight, C1PW0, will be appropriate if only parental are analyzed. I use this weight to calculate gaps in family background and investment. I find that the estimates are close in most cases.

[9]For the cognitive factor, the factor loadings are much higher for test scores than for teacher ratings. For this reason, I predict factor scores using only the test scores. An alternative prediction method calculates the score by weighting measures such that resulting noise in the factor score is minimized. I have also run analyses using this method, with similar results.

[10]In earlier versions of this paper, I presented racial gaps in the normalized measure.

Given the kindergarten skills and investment, how much worse than expected do blacks do? This graph starts at zero, and shows that blacks fall behind by approximately .3 sd. In contrast, Hispanics stay close to expectations. While these results are strong for blacks, the fall-off is well-documented (Fryer and Levitt, 2006).

9.4 Noncognitive Skills

The results are given in figure 11 and table 23. Asians stand out. In raw gaps, they start with an advantage of .16 sd, which increases to more than .5 sd by 5th grade. These advantages are even increased when controlling for investments: Since Asians lag in investment, their gaps now rise to more than .6 sd. Not surprisingly, they deviate strongly from expectations, at more than .6 sd.

In contrast, gaps for other groups are much more modest: Blacks start out with a lag of about .3 sd, which drops strongly by third grade, but recovers by grade 5. The raw skill gaps of latinos are close to zero. The second panel controls for investment and family background. Since blacks score lower on both factors, controlling for these factors reduces the skill gaps for blacks and latinos by more than .15 sd. The third panel provides the deviations from expected gaps: Blacks decrease by .15 over the sample period and Hispanics exceed expectations by .25 sd.

10 Effect of Skills On Economic Outcomes

The ECLS-K allows to document racial skill gaps in childhood, but it cannot be used to forecast how these skills affect economic outcomes. To overcome this, I use data from the British Cohort Study of 1970 (BCS70 from now on), which contains childhood skill measures and long-term follow-ups.

10.1 Data Description

The BCS70 started as a survey of all $17,198$ babies born in England, Scotland, Wales, and Northern Ireland in the week of 5-11 April 1970 (Bynner et al., 2000). Since then, follow-ups were obtained at age 5,10,16,26,30,34, and 38. The surveys are well-suited to understand the development and impact of child skills: In all childhood waves, the data contains extensive information on test scores and behavior. Fortunately, the age-10 survey contains a wide range of behavior ratings by teachers. This allows linking the ECLS-K and the BCS, because the ECLS-K contains teacher-rated skill gaps by 5th grade, when children are age 10-11.

10.2 Skill Measures

The teacher ratings in the BCS70 and the ECLS-K are not identical. For this reason, I seek to match the cognitive and noncognitive skill factor as well as possible. To proxy the cognitive factor, I take an average of two tests provided in the BCS70

- Math: Friendly Math Test (FMT), sum of correct answers.
- Reading: Edinburgh Reading Test (ERT), sum of correct answers.

The noncognitive measures are proxied as averages of variable responses. Teachers rank students between 1 and 100 on each of these questions.

- Self-control: Displays outbursts of temper, pays attention in class, completes tasks
- Interpersonal: Has many friends
- Externalizing: Quarrels with other kids, Destroys belongings, Excitable or Impulsive, Bullies other children
- Internalizing: Afraid of new things, Behaves nervously, Rather solitary, Sullen or Sulky

I then take an average of the four components (self-control, interpersonal, externalizing, internalizing) to arrive at the noncognitive score.

10.3 Results

This subsection provides regression results, using skills at age 10 are to predict economic outcomes. The outcome variables are:

- Years of education completed by age 38
- Log wage by age 30, 34, and 38. Age 38 is the age at the most recent survey.

The control variables are:

- Parental background: Years of education of mother and father, presence of father in household, family status, mother's age at birth.
- Dummy for region of birth
- Female dummy

The results are presented in table 27. This table makes it clear that test scores are more predictive for economic outcomes than the noncognitive measure. A one sd increase in test scores is associated with 1.3 more years of education. Noncognitive skills are associated with .22 more years.

Measuring the impact of skills on wages is less straightforward. Optimally, one would want to know how skills affect lifetime income. Income early in life is a potentially poor proxy. For instance, since skills affect education, and highly educated individuals have higher wage increases over time, the coefficient on skills is potentially downward-biased using early-life wages. This can be seen in the regression results: When going from wages at age 30 to wages at age 38, the coefficients on both skills increase: From .11 to .15 for test scores, and from .03 to .04 for noncognitive skills. In the remainder, I forecast using the production function at age 38. Results are similar when instead using an average of all regression coefficients.

How much do skills contribute to racial gaps in outcomes? To predict this, I combine the racial gaps from the ECLS-K with the coefficients obtained from the BCS70. The racial gaps are obtained from a regression using the ECLS-K that controls for parental background (Mother's years of education, mother's age at birth, log family income, presence of father) and racial dummies. This forecasting exercise requires the assumption that the production function is identical for these two very different samples. Results are presented in table 28. The top panel describes the estimated skill gaps, the bottom panel the predicted outcome gaps.

The skill gaps differ from the gaps calculated in the skill formation model, because they control for covariates, and do not condition on initial skill. For instance, the estimated cognitive gap of Hispanics stands at .25, while the raw gaps stands at .72. This results because Hispanics lag strongly in the parental background characteristics that predict test scores. The strong effect of control variables suggest that differing specifications (for instance, using parental background before Kindergarten) would reduce the gaps even further. However, this is not likely a problem for Asians: Their parental background is similar to whites, and controlling for it changes gaps only negligibly.

For both blacks and Hispanics, the models predict large gaps in educational attainment and log wages, with both gaps being driven by cognitive skills. Asians are estimated without test score gaps, but with a .5 advantage over whites in noncognitive skills. This translates into an *advantage* of 2.5 log points due to noncognitive skills.

Noncognitive skills can explain between 8% and 16% of the log wage gaps given in table 25. While this range seems small, it is likely an underestimate:

- The estimation uses a single measure of noncognitive skills, provided by a single teacher at a single point in time. Since noncognitive skills are multi-faceted, it is likely that this single measure underestimates the impact of noncognitive skills.

- It is likely that noncognitive skills are more malleable than cognitive skills. In that case, a measure taken later during childhood will potentially predict outcomes better.
- Wage gaps for Asians are likely smaller once accounting for the strong locational sorting of Asians. As presented above, a control for state of residence reduces gaps by approximately 50%.
- If there are racial peer effects of skills, individual gaps under-predict group gaps in outcomes.

Effects on Crime/Arrests Section 3 presented evidence that Asians not only have better economic outcomes than whites, but also that they are much less likely to be arrested. Using the NELS-88 data, it is possible to establish that noncognitive skills do affect (self-reported) arrest rates. Here, the outcomes and measures are too different to attempt a quantitative forecast, but it is possible to show that noncognitive skills strongly affect arrest rates. To do so, I run a probit regression: A dummy for being arrested (by the respondents' early 20s) is regressed on family background and skills by 8th grade. I calculate a cognitive measure using test scores, and a noncognitive measure using teacher ratings of behavior in class: Being disruptive (externalizing), passive (internalizing), and attentive (self-control). Using a linear probability model, I find that both noncognitive and cognitive skills predict lower arrest rates. The explanatory power of noncognitive skills is stronger than the explanatory power of test scores (which is borderline significant).

11 Conclusion

Recent research has highlighted that a wide range of noncognitive skills can be as predictive for life outcomes as cognitive skills such as IQ or test scores (Almlund, Duckworth, Heckman, and Kautz, 2011). This suggests that racial gaps in outcomes can be partially explained via gaps in noncognitive skills. This paper shows that there is one salient feature: Asians show a .5 sd advantage over Whites in teacher-rated noncognitive skills. Using an external sample, this gap can be forecast to explain a 2.5 log point difference in wages.

For estimation, this paper uses a dynamic factor model of skill formation that incorporates the role of family background and investments. The model can identify counterfactual racial skill gaps, equating investments between groups. It is well known that Blacks underperform in cognitive skills (here measured via test scores). For teacher ratings of noncognitive skills, Asians have high raw gaps, which cannot be explained by racial gaps in (observable) inputs.

There are many questions that this paper does not answer. It does not answer whether it is possible to bridge the racial gaps in skills using public policy. Consider Asians' noncognitive skills: Are they high because Asian parents have a parenting behavior that increases noncognitive skills, and this behavior is not captured using the skill formation model?[11] Or are they high due to racial peer effects? It is also possible that racial groups differ in heritable components that affect skills. This paper does not allow to distinguish between these explanations.

This paper highlights that, for noncognitive skills, it matters whether one asks parents or teachers. For instance, by first grade, Asians lag strongly in social behavior when using parental responses. In contrast, teachers rank them ahead of whites. An intriguing interpretation is that Asians parents have very high expectations of their children, and judge them to fail according to expectations. Potentially, these expectations play a role in explaining high skills. Alternatively, parents and teachers report accurately, but Asian children behave differently in the home and in school. Again, an answer to this complicated question is left to future work.

Noncognitive skills can help in understanding a wide range of phenomena commonly referred to as culture: For instance, a standard explanation for Asians' high level of education is that Asians have "cultural values that promote educational endeavors" (Sue and Okazaki, 1990). Assume, for the sake of simplicity, that childrens' cognitive skills are completely unaffected by inputs or environment. In

[11] This result is robust to using additional measures of parental investment. In earlier versions of this paper, I have used a large range of variables to explain noncognitive skills, which affects the results only little.

that case, it seems natural that Asian culture translates into noncognitive skills (such as self-control) that increase educational attainment. Thus, measurable noncognitive skills allow to operationalize the concept of culture.

References

M. Almlund, A.L. Duckworth, J.J. Heckman, and T.D. Kautz. Personality psychology and economics. Technical report, National Bureau of Economic Research, 2011.

J.G. Altonji and R.M. Blank. Race and gender in the labor market. *Handbook of labor economics*, 3: 3143–3259, 1999.

J. Bynner, N. Butler, E. Ferri, P. Shepherd, and K. Smith. The design and conduct of the 1999–2000 surveys of the national child development study and the 1970 british cohort study. *Centre for Longitudinal Studies, Cohort Studies, Working Paper*, (1), 2000.

A. Claessens, G. Duncan, and M. Engel. Kindergarten skills and fifth-grade achievement: Evidence from the ecls-k. *Economics of Education Review*, 28(4):415–427, 2009.

F. Cunha and J.J. Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782, 2008.

T.A. DiPrete and J.L. Jennings. Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, 2011.

J. Fitzgerald, P. Gottschalk, and R. Moffitt. An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, 33(2):251–299, 1998.

R.G. Fryer and S.D. Levitt. The black-white test score gap through third grade. *American Law and Economics Review*, 8(2):249, 2006.

E.A. Hanushek. Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis*, 19(2):141, 1997.

J.J. Heckman and E.J. Vytlacil. Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics*, 6, 2007.

M. Najarian, J.M. Pollack, and A.G. Sorongon. Early Childhood Longritudinal Study, Kindergarten Class of 1998-1999 (ECLS-K). Psychometric Report for the Eighth Grade (NCES 2009-002). Technical report.

D.A. Neal and W.R. Johnson. The role of pre-market factors in black-white wage differences. Technical report, National Bureau of Economic Research, 1995.

A. Sakamoto, K.A. Goyette, and C.H. Kim. Socioeconomic attainments of asian americans. *Annual Review of Sociology*, 35:255–276, 2009.

S. Sue and S. Okazaki. Asian-american educational achievements: A phenomenon in search of an explanation. *American Psychologist; American Psychologist*, 45(8):913, 1990.

P.E. Todd and K.I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):3–33, 2003.

P.E. Todd and K.I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, 1(1):91–136, 2007.

K. Tourangeau, C. Nord, Thanh Le, J.M. Pollack, S. Atkins-Burnett, and E. Germino Hausken. Early Childhood Longritudinal Study, Kindergarten Class of 1998-1999 (ECLS-K). Combined Users Manual for the ECLS-K Fifth-Grade Data Files and Electronic Codebooks. Technical report, a.

K. Tourangeau, C. Nord, L. Thanh, and A. G. Sorongon. Early Childhood Longritudinal Study, Kindergarten Class of 1998-1999 (ECLS-K). Combined User's Manual. Technical report, b.

K. Tourangeau, C. Nord, T. Lê, and A. Sorongon. Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), Combined Users Manual for the ECLS-K Eigth-Grade Data Files and Electronic Codebooks (NCES 2006-032). Technical report, 2009.

A Tables and Figures

A.1 Figures

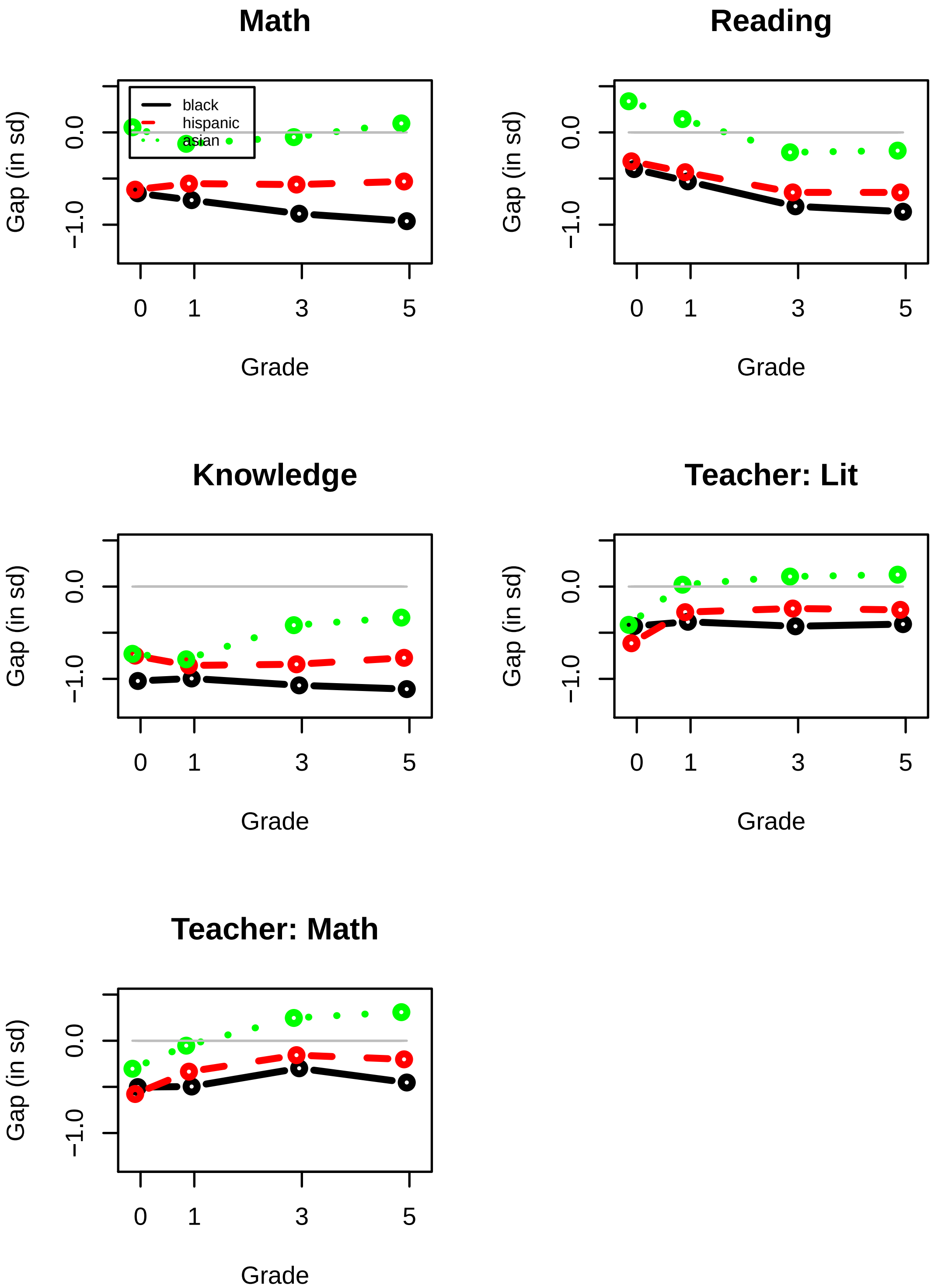

Figure 3: Gaps: Cognitive skills

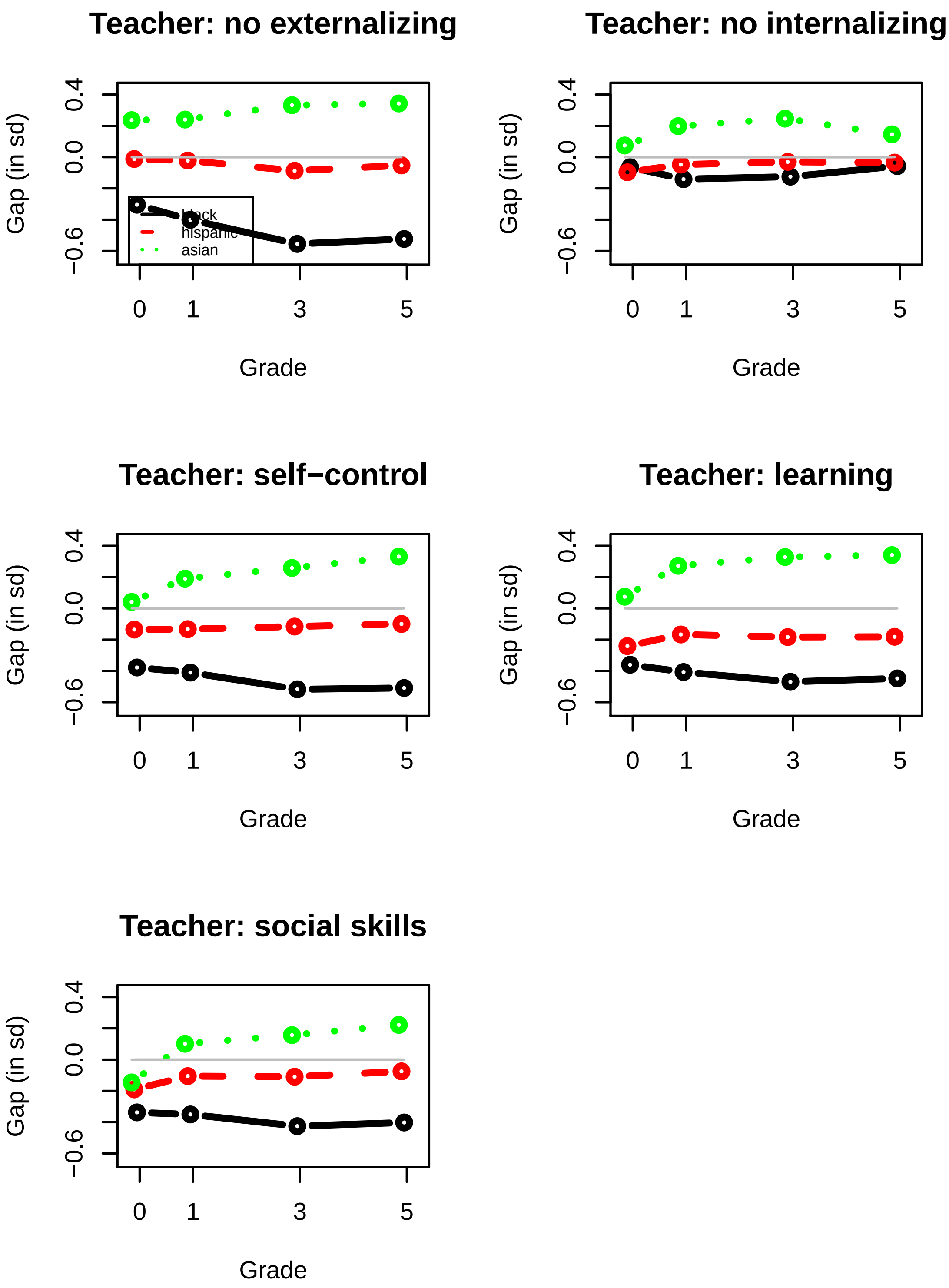

Figure 4: Gaps: Noncognitive, teacher reports

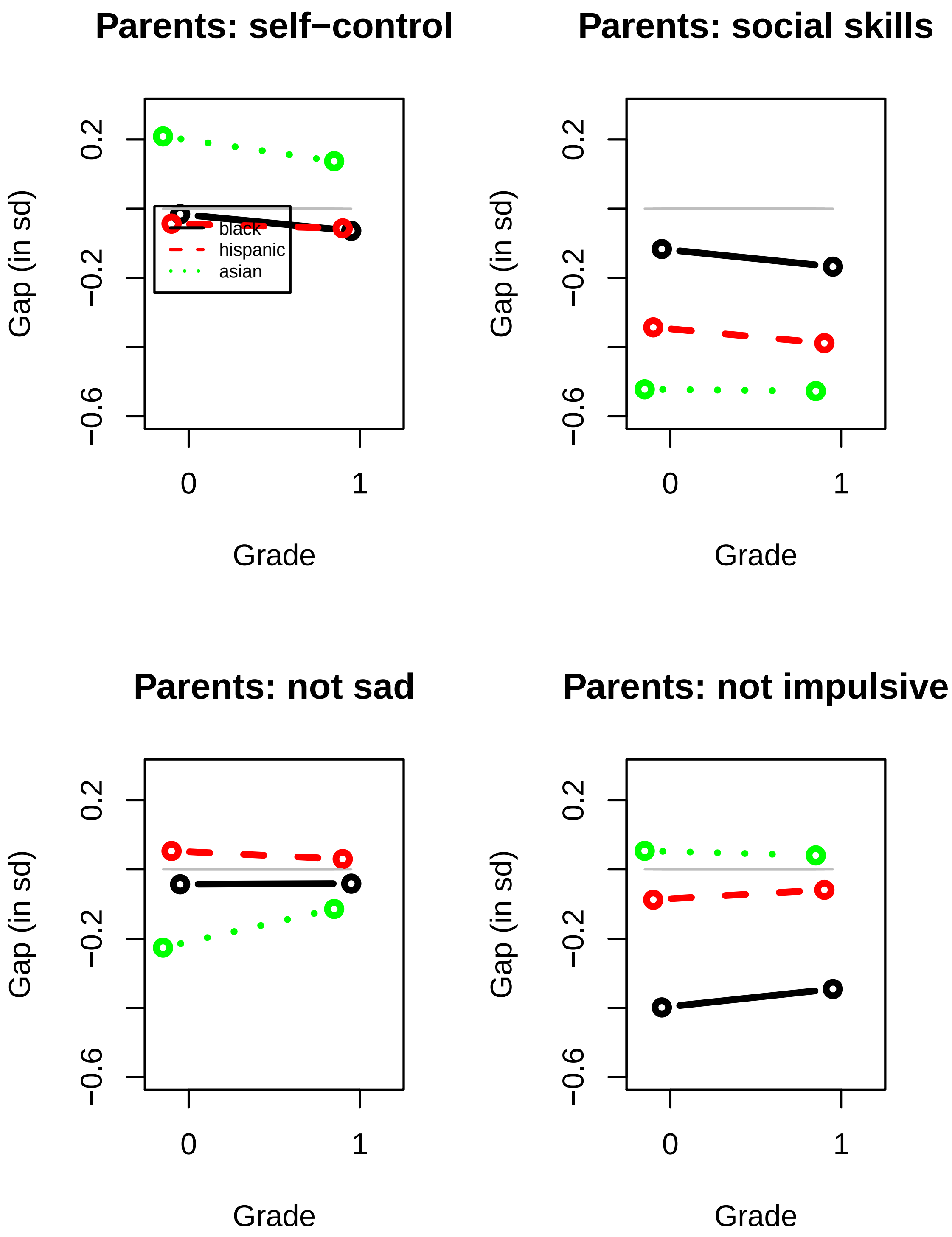

Figure 5: Gaps: Noncognitive, parental measures

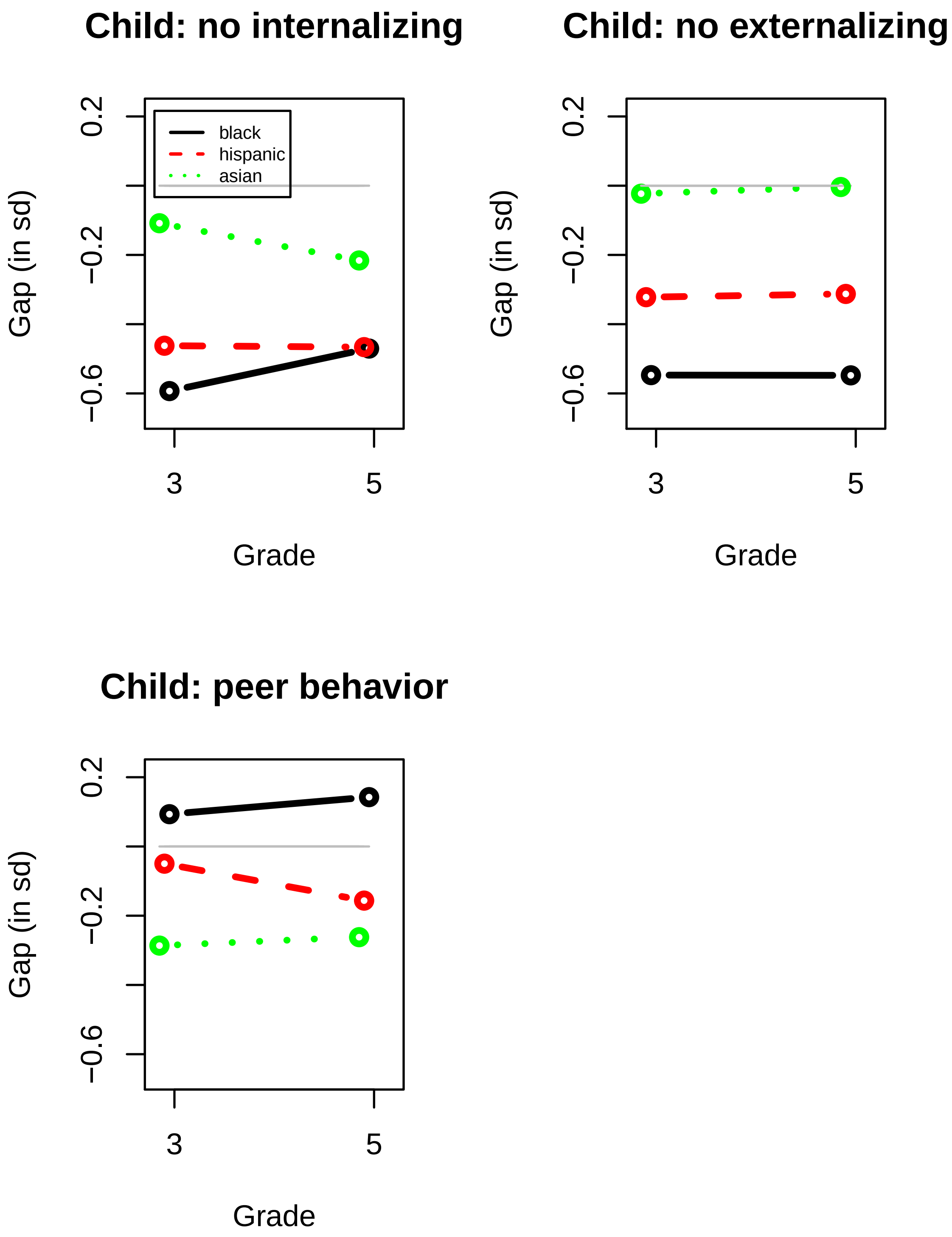

Figure 6: Gaps: Noncognitive, self reports

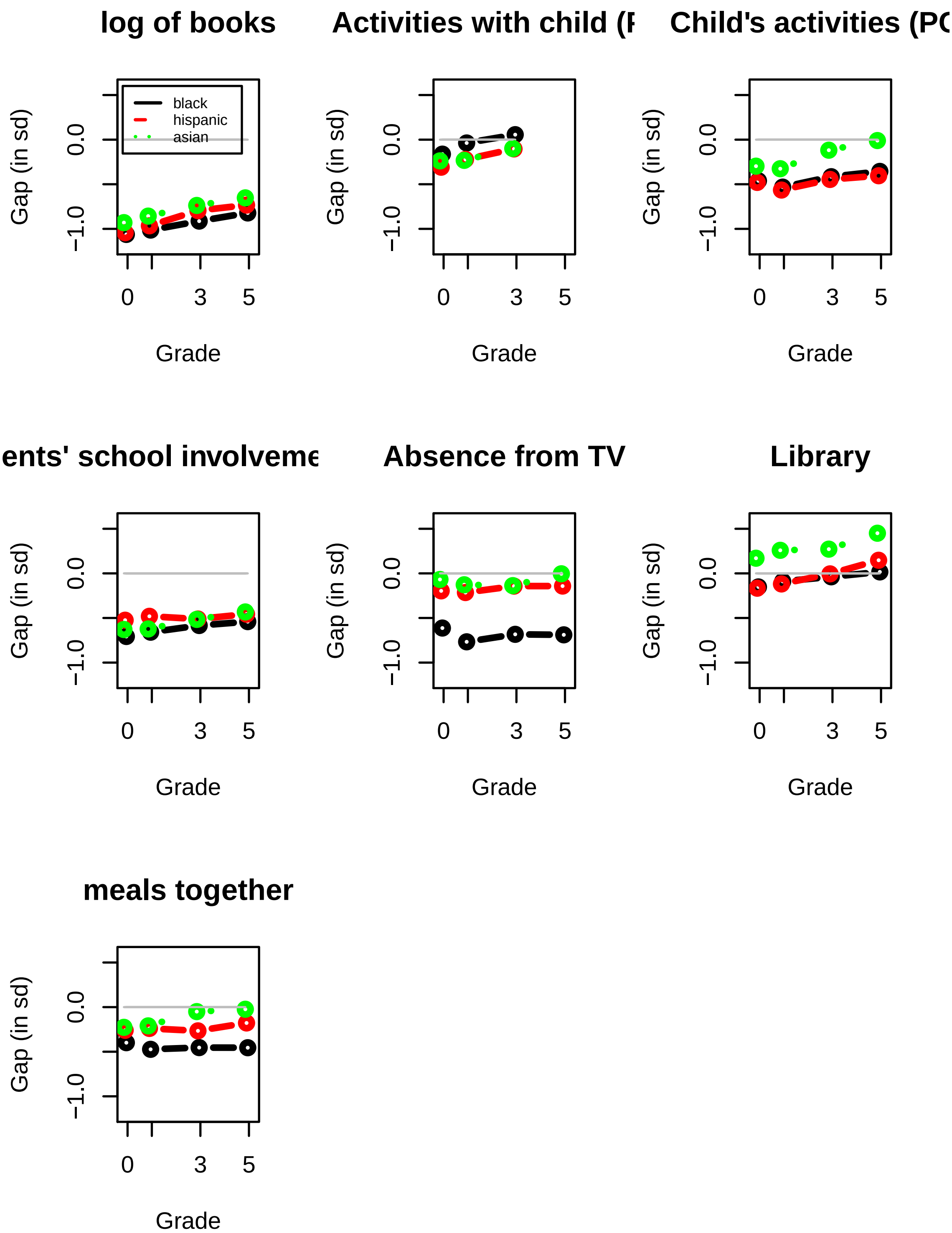

Figure 7: Gaps: Parental investment

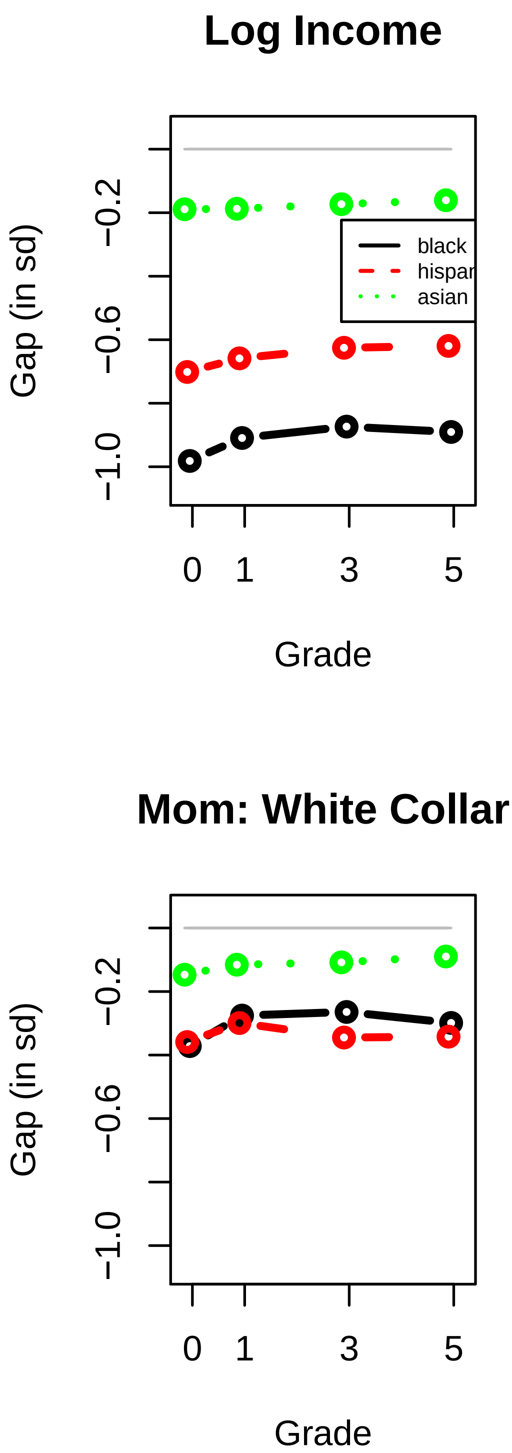

Figure 8: Gaps: Parental background

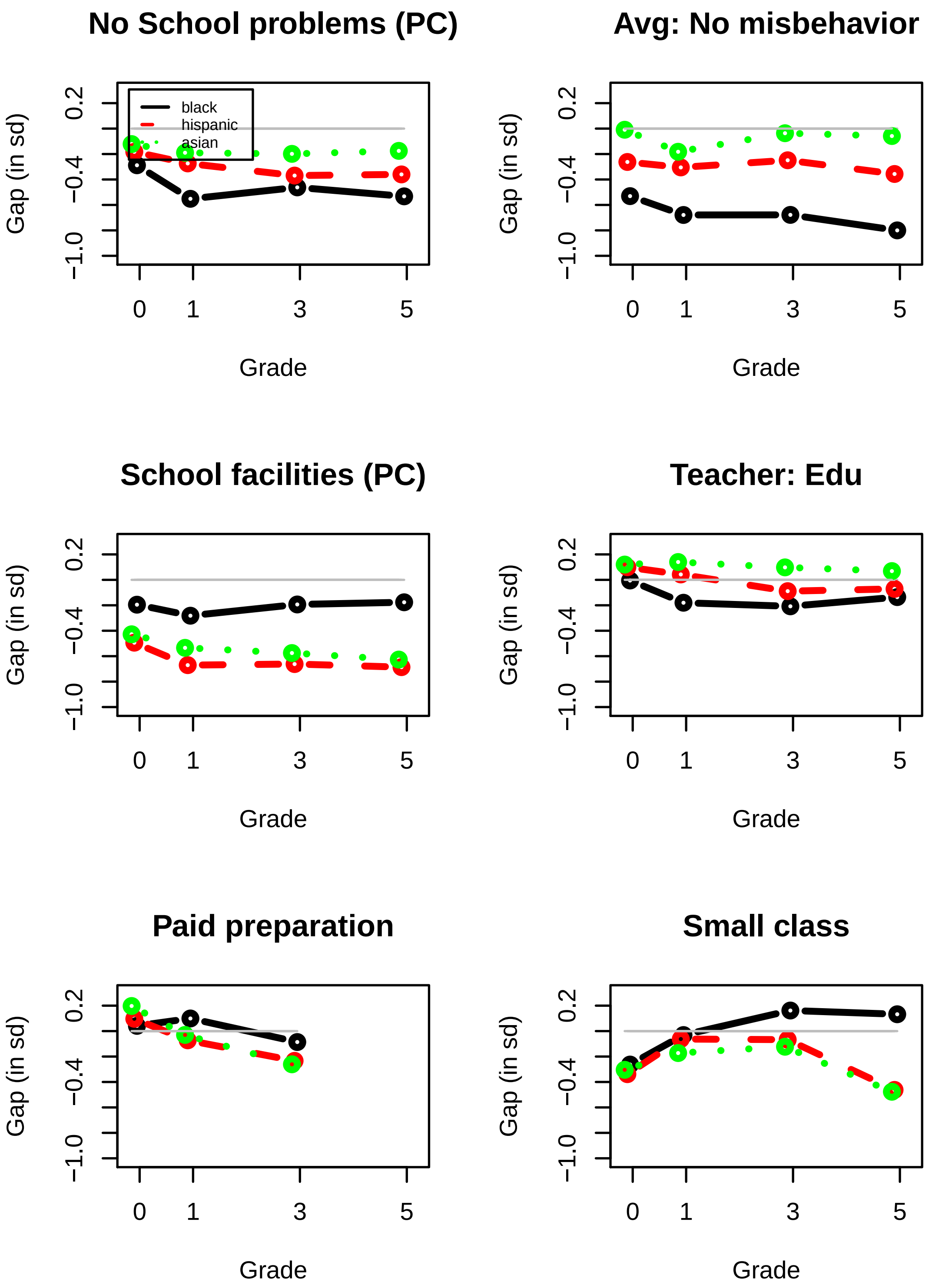

Figure 9: Gaps: School quality

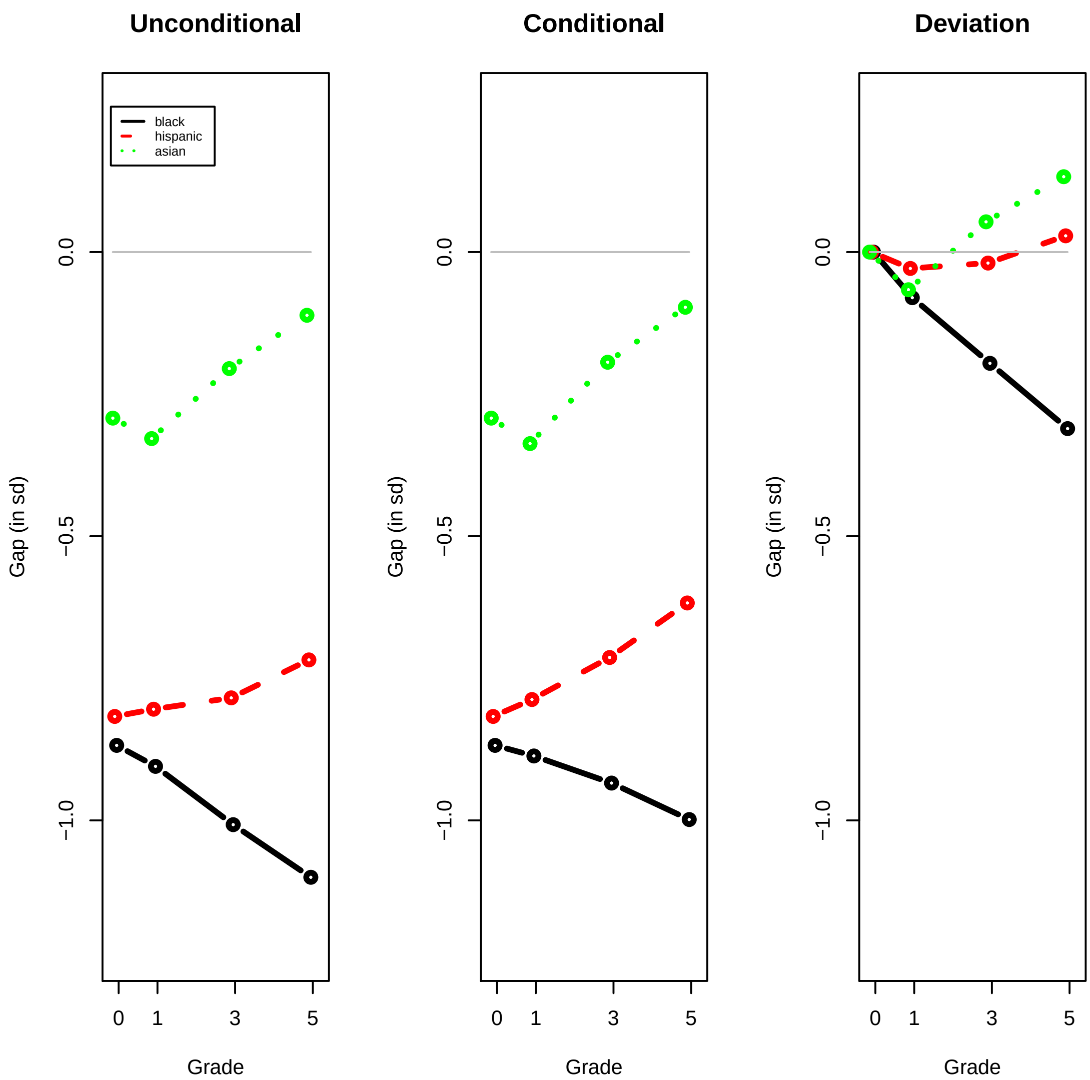

Figure 10: Counterfactual gaps: Cognitive factor

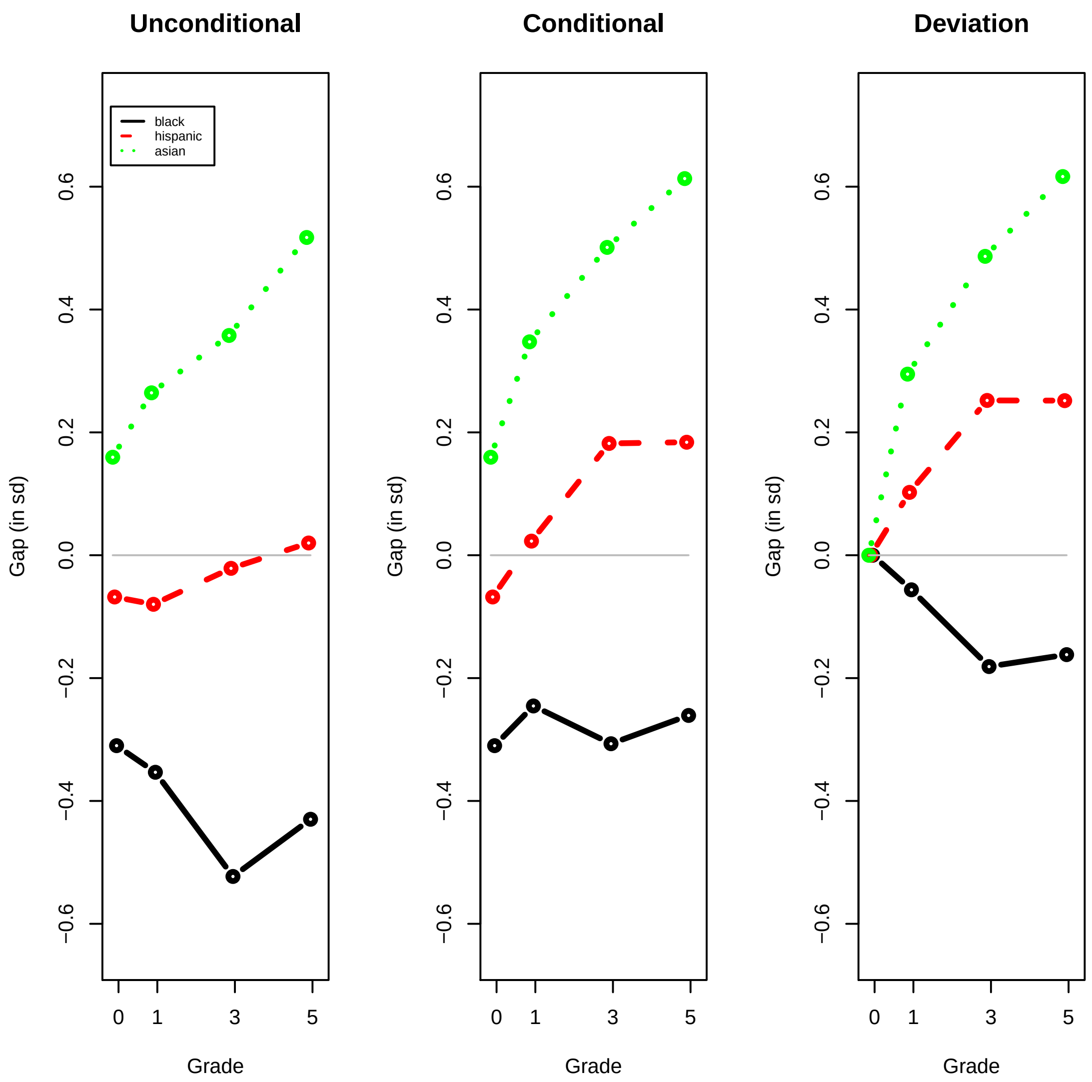

Figure 11: Counterfactual gaps: Noncognitive factor (teacher rating)

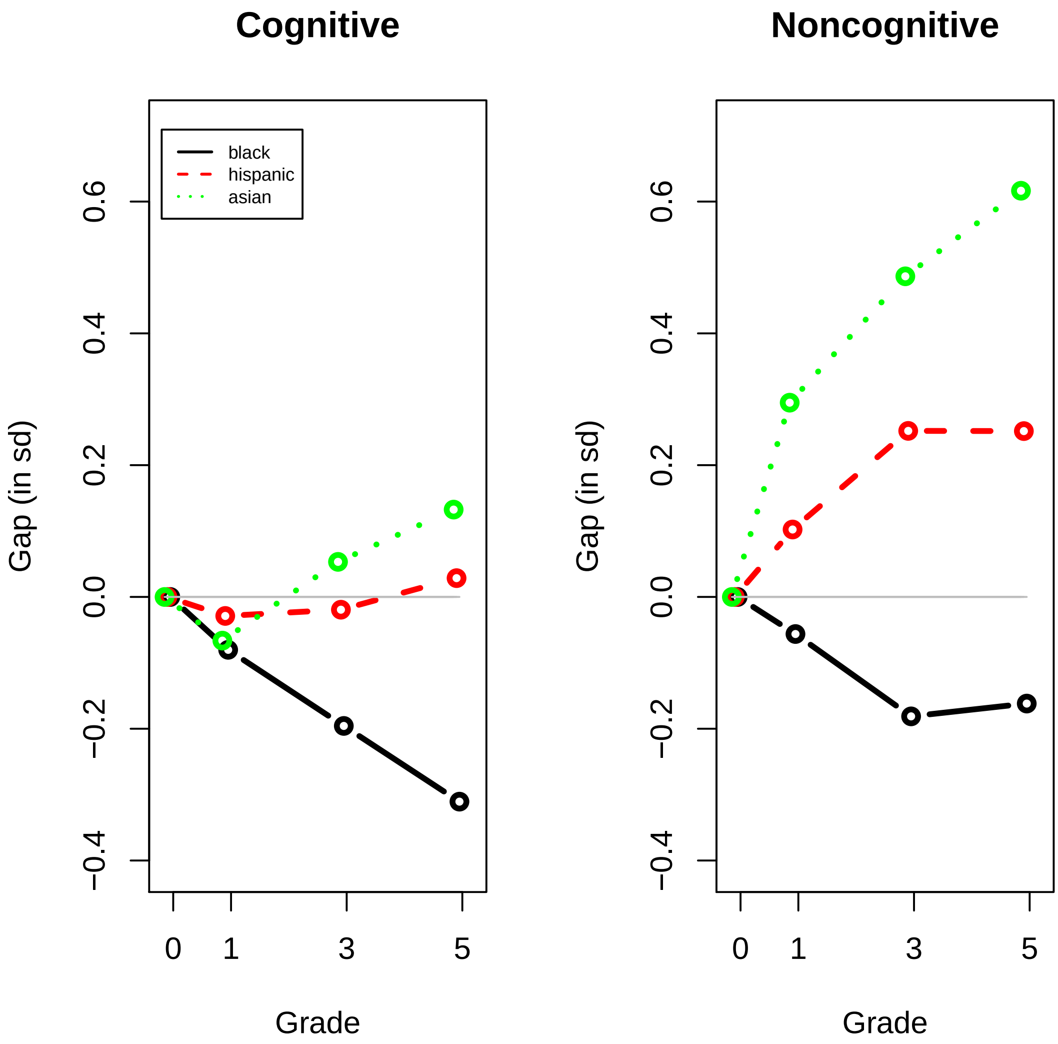

Figure 12: Skill Gaps, deviation from expectation

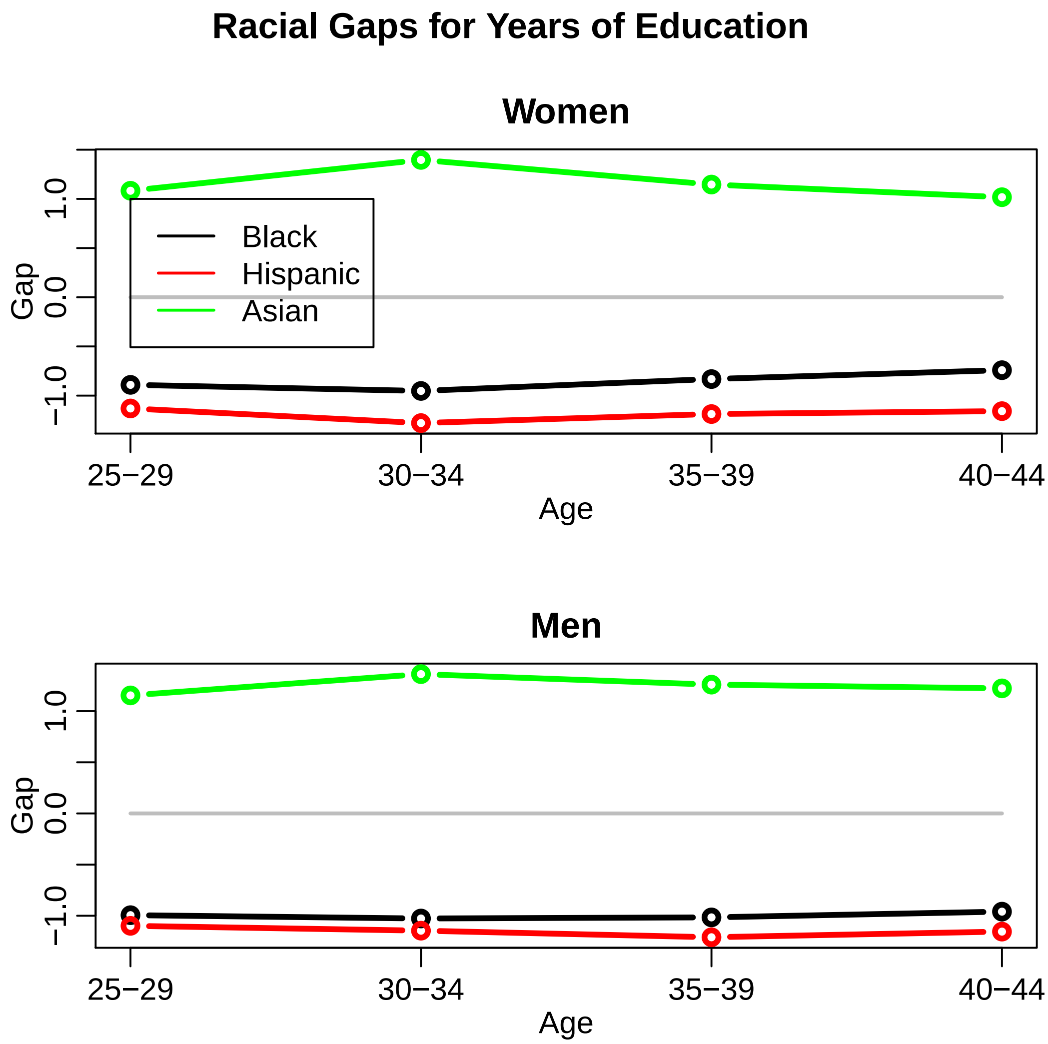

Figure 13: Racial Gaps in Years Of Education, controlling for year dummies. Source: ACS 2006-2010. Universe: US-born respondents

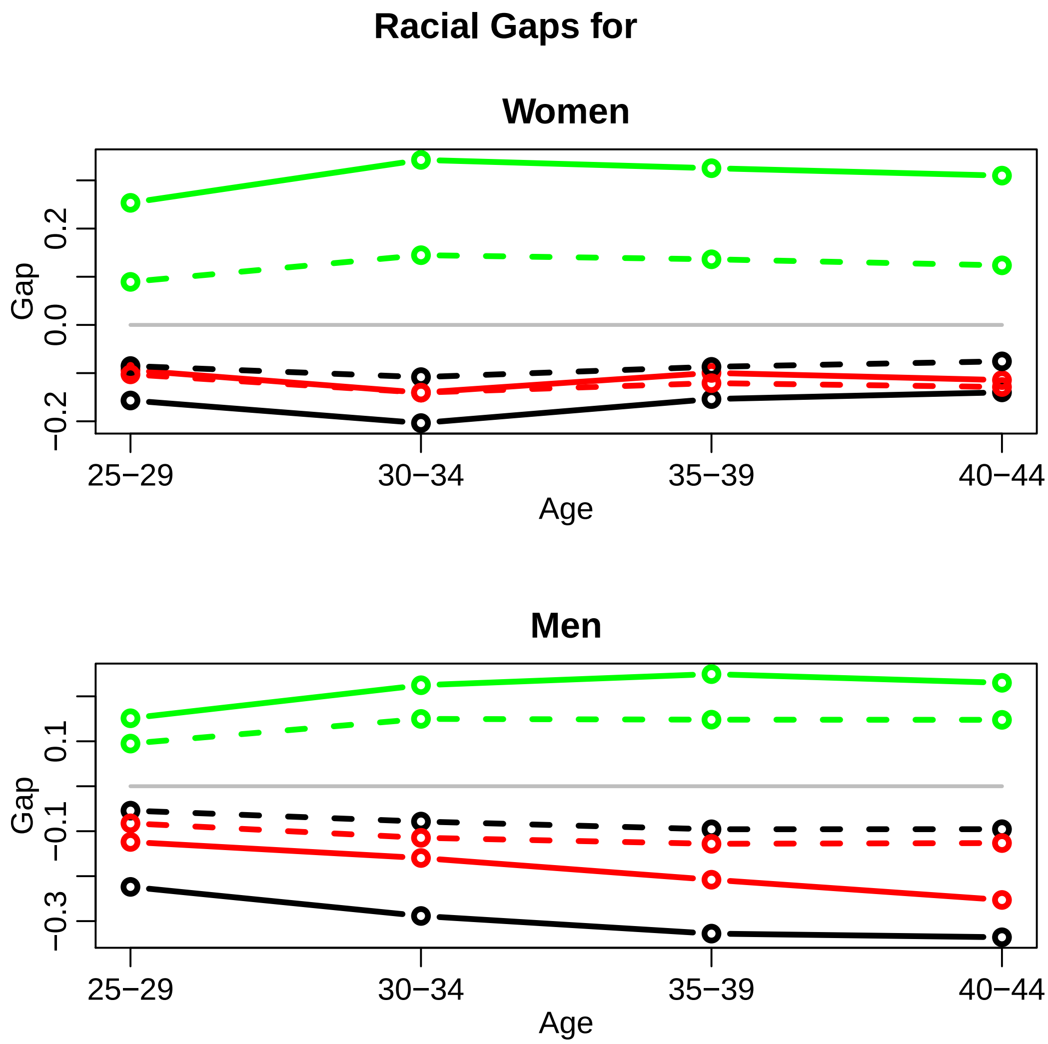

Figure 14: Racial Gaps in Log Hourly Wage, controlling for year dummies. Solid line gives raw gaps, dashed lines show the gap that is explained by a linear control for years of education. Source: ACS 2006-2010. Universe: US-born respondents, not in school

A.2 Tables

Measures of skills: x *denotes presence*	Wave			
Measure	K	1	3	5
Score in math test	x	x	x	x
Score in reading test	x	x	x	x
Teacher evaluation: math	x	x	x	x
Teacher evaluation: reading	x	x	x	x
Teacher evaluation: learning	x	x	x	x
Teacher evaluation: self-control	x	x	x	x
Teacher evaluation: interpersonal skills	x	x	x	x
Teacher evaluation: externalizing	x	x	x	x
Teacher evaluation: internalizing	x	x	x	x
Self-description: externalizing			x	x
Self-description: internalizing			x	x
Self-description: peer behavior			x	x
Parent evaluation: self-control	x	x		
Parent evaluation: social skills	x	x		
Parent evaluation: sad	x	x		
Parent evaluation: impulsive	x	x		

Table 5: Measures of skills

Measures of PI: *x denotes presence*	Grade			
Measure	K	1	3	5
Number of meals together		x	x	
Number of library visits		x	x	
Hours watched TV per day		x	x	
Child participation index		x	x	
Activities index		x	x	
Tells stories	x	x	x	
Sings songs	x	x	x	
Help with arts and crafts	x	x	x	
Involve child in household chores	x	x	x	
Play games	x	x	x	
Talk about nature	x	x	x	
Build something with child	x	x	x	
Play sports together	x	x	x	
Practice reading, writing or numbers		x	x	
Read books	x	x	x	
Number of books	x	x	x	x
Involvement index	x	x	x	x
Involvement in School	x	x	x	x
Attended open house	x	x	x	x
Gone to PTA conference	x	x	x	x
Attended school or class event	x	x	x	x
Acted as volunteer	x	x	x	x
Participated in fundraising	x	x	x	x

Table 6: Measures of parental investment

Measures of school quality: *x denotes presence*	Wave			
Measure	K	1	3	5
School Avg: misbehavior (teacher-rated)	x	x	x	x
School Avg: school spirit (teacher-rated)	x	x	x	x
School Avg: students capable (teacher-rated)	x	x	x	x
School Avg: Teacher's motivation	x	x	x	x
School crime problems index	x	x	x	x
Facilities index	x	x	x	x
Class size	x	x	x	x
Teacher's years of education	x	x	x	x
Teacher's work hours	x	x	x	x

Table 7: Measures of school quality

	Cognitive K	NC Teacher K	NC Parent K	Cognitive 1	NC Teacher 1	NC Parent 1
Math	0.8 (0)			0.73 (0)		
Reading	0.73 (0.017)			0.8 (0.019)		
Knowledge	0.63 (0.018)			0.62 (0.019)		
Teacher: Lit	0.67 (0.021)	0.23 (0.019)	-0.17 (0.028)	0.81 (0.024)	0.21 (0.017)	-0.18 (0.028)
Teacher: Math	0.6 (0.022)	0.24 (0.02)	-0.18 (0.029)	0.75 (0.024)	0.2 (0.018)	-0.19 (0.029)
Teacher: no internalizing	0.15 (0.019)	0.35 (0.021)	-0.01 (0.022)	0.21 (0.019)	0.3 (0.02)	0.01 (0.022)
Teacher: learning	0.35 (0.016)	0.72 (0.02)	-0.13 (0.021)	0.42 (0.017)	0.65 (0.018)	-0.11 (0.022)
Teacher: social skills	0.1 (0.013)	0.85 (0.021)	-0.06 (0.017)	0.06 (0.013)	0.87 (0.019)	-0.02 (0.015)
Parents: self-control	0.09 (0.03)	0 (0.023)	0.68 (0.039)	-0.01 (0.031)	0.04 (0.029)	0.65 (0.038)
Parents: social skills	0.08 (0.021)	0 (0.021)	0.21 (0.024)	0.05 (0.022)	0.02 (0.022)	0.28 (0.025)
Parents: not impulsive	0.12 (0.025)	0.09 (0.022)	0.5 (0.031)	-0.01 (0.027)	0.12 (0.026)	0.53 (0.033)
Teacher: no externalizing		0.7 (0)			0.76 (0)	
Teacher: self-control		0.89 (0.02)			0.9 (0.017)	
Parents: not sad			0.46 (0)			0.45 (0)

Table 8: Factor loadings for skill factors when including all measures. Wave K-1. Standard errors in parentheses.

	Cognitive 3	NC Teacher 3	NC Child 3	Cognitive 5	NC Teacher 5	NC Child 5
Math	0.86 (0)			0.85 (0)		
Reading	0.87 (0.013)			0.87 (0.014)		
Knowledge	0.8 (0.014)			0.81 (0.014)		
Teacher: Lit	0.7 (0.019)	0.23 (0.015)	-0.1 (0.021)	0.6 (0.017)	0.28 (0.015)	0.01 (0.019)
Teacher: Math	0.65 (0.02)	0.18 (0.017)	-0.11 (0.023)	0.62 (0.018)	0.15 (0.015)	0.01 (0.019)
Teacher: no internalizing	0.17 (0.021)	0.37 (0.019)	-0.01 (0.025)	0.14 (0.02)	0.36 (0.018)	0.03 (0.023)
Teacher: learning	0.32 (0.016)	0.72 (0.017)	-0.03 (0.018)	0.25 (0.014)	0.74 (0.017)	0.02 (0.016)
Teacher: social skills	0.07 (0.014)	0.9 (0.018)	-0.08 (0.017)	0.04 (0.013)	0.88 (0.018)	-0.05 (0.015)
Child: no externalizing	-0.17 (0.026)	0.12 (0.019)	0.84 (0.041)	-0.05 (0.02)	0.28 (0.017)	0.7 (0.033)
Child: peer behavior	-0.29 (0.023)	0.1 (0.02)	0.34 (0.027)	-0.14 (0.021)	0.14 (0.019)	0.35 (0.025)
Teacher: no externalizing		0.76 (0)			0.75 (0)	
Teacher: self-control		0.89 (0.016)			0.89 (0.017)	
Child: no internalizing			0.74 (0)			0.73 (0)

Table 9: Factor loadings for skill factors when including all measures. Wave 3-5. Standard errors in parentheses.

	Cognitive K	NC Teacher K	Cognitive 1	NC Teacher 1	Cognitive 3	NC Teacher 3	Cognitive 5	NC Teacher 5
Math	0.84 (0)		0.82 (0)		0.83 (0)		0.84 (0)	
Reading	0.73 (0.021)		0.83 (0.02)		0.84 (0.019)		0.84 (0.019)	
Knowledge	0.7 (0.021)		0.73 (0.021)		0.76 (0.02)		0.77 (0.019)	
Teacher: Lit	0.58 (0.022)	0.21 (0.022)	0.67 (0.021)	0.22 (0.018)	0.65 (0.02)	0.23 (0.019)	0.61 (0.02)	0.29 (0.019)
Teacher: Math	0.5 (0.023)	0.24 (0.023)	0.62 (0.021)	0.19 (0.02)	0.62 (0.021)	0.15 (0.02)	0.64 (0.021)	0.11 (0.02)
Teacher: no internalizing	0.12 (0.024)	0.36 (0.026)	0.21 (0.024)	0.31 (0.024)	0.2 (0.023)	0.36 (0.024)	0.16 (0.023)	0.39 (0.024)
Teacher: learning	0.26 (0.018)	0.73 (0.025)	0.34 (0.018)	0.66 (0.019)	0.29 (0.017)	0.71 (0.019)	0.26 (0.017)	0.73 (0.019)
Teacher: social skills	0.06 (0.017)	0.85 (0.027)	0.03 (0.015)	0.87 (0.019)	0.03 (0.015)	0.88 (0.019)	0.04 (0.015)	0.87 (0.019)
Teacher: no externalizing		0.69 (0)		0.75 (0)		0.77 (0)		0.77 (0)
Teacher: self-control		0.89 (0.026)		0.9 (0.018)		0.9 (0.018)		0.9 (0.018)

Table 10: Factor loadings for skill factors when using just teacher measures. Waves K-5. Standard errors in parentheses.

K				
	Cognitive	NC Teacher	NC Parent	
Cognitive	1			
NC Teacher	0.13	1		
NC Parent	0.2	0.43	1	
1				
	Cognitive	NC Teacher	NC Parent	
Cognitive	1			
NC Teacher	0.22	1		
NC Parent	0.32	0.24	1	
3				
	Cognitive	NC Teacher		NC Child
Cognitive	1			
NC Teacher	0.24	1		
NC Child	0.58	0.41		1
5				
	Cognitive	NC Teacher		NC Child
Cognitive	1			
NC Teacher	0.21	1		
NC Child	0.44	0.24		1

Table 11: Factor correlations for skill factors. Waves K-5.

	Investment K	Investment 1	Investment 3	Investment 5
log of books	0.55 (0)	0.53 (0)	0.45 (0)	0.44 (0)
Activities with child (PC)	0.32 (0.045)	0.2 (0.042)	0.11 (0.042)	
Child's activities (PC)	0.47 (0.05)	0.53 (0.053)	0.5 (0.061)	0.48 (0.065)
Parents' school involvement (PC)	0.6 (0.056)	0.54 (0.056)	0.5 (0.063)	0.54 (0.07)
Absence from TV	0.3 (0.045)	0.41 (0.048)	0.4 (0.054)	0.41 (0.057)
Library	0.32 (0.045)	0.35 (0.046)	0.25 (0.047)	0.25 (0.049)
meals together	0.31 (0.045)	0.33 (0.045)	0.31 (0.049)	0.25 (0.048)

Table 12: Factor loadings for investment. Waves K-5. Standard errors in parentheses.

	Background K	Background 1	Background 3	Background 5
Log Income	0.66 (0)	0.68 (0)	0.69 (0)	0.67 (0)
SES	0.97 (0.043)	0.98 (0.04)	0.98 (0.035)	0.98 (0.042)
Mom: White Collar	0.55 (0.039)	0.55 (0.037)	0.54 (0.035)	0.54 (0.037)

Table 13: Factor loadings for family background. Waves K-5. Standard errors in parentheses.

	Quality K	Quality 1	Quality 3	Quality 5
No School problems (PC)	0.62 (0)	0.62 (0)	0.6 (0)	0.57 (0)
Avg: No misbehavior	0.16 (0.048)	0.08 (0.048)	0.2 (0.057)	0.31 (0.045)
School facilities (PC)	0.1 (0.044)	0.03 (0.044)	-0.03 (0.045)	0 (0.044)
Teacher: Edu	-0.03 (0.044)	0.05 (0.047)	-0.14 (0.046)	-0.06 (0.043)
Paid preparation	-0.09 (0.043)	-0.17 (0.047)	0 (0.043)	
Small class	0 (0.046)	-0.18 (0.062)	-0.12 (0.062)	-0.07 (0.048)

Table 14: Factor loadings for school quality. Waves K-5. Standard errors in parentheses.

	Quality K	Quality 1	Quality 3	Quality 5
Avg: No misbehavior	0.26 (0)	0.26 (0)	0.29 (0)	0.35 (0)
No School problems (PC)	0.53 (0.042)	0.52 (0.037)	0.5 (0.041)	0.52 (0.033)
School facilities (PC)	0.12 (0.02)	0.1 (0.02)	0.04 (0.019)	0.02 (0.019)

Table 15: Factor loadings for school quality. Waves K-5. Standard errors in parentheses.

	Cognitive			NC Teacher		
Grade	1	3	5	1	3	5
Lagged: Cognitive	0.971	0.969	0.968	0.109	0.105	0.073
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Lagged: NC Teacher	-0.015	0.001	0.012	0.468	0.529	0.538
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)

Table 16: Production function. Waves K-5. Standard errors in parentheses. Specification without investment.

	Cognitive			NC Teacher		
Grade	1	3	5	1	3	5
Lagged: Cognitive	0.936	0.939	0.943	0.114	0.116	0.085
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Lagged: NC Teacher	0.013	0.014	0.013	0.446	0.507	0.527
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)

Table 17: Production function. Waves K-5. Standard errors in parentheses. Specification uses without investment. Allowing for autocorrelated measurement errors.

Grade	Cognitive 1	3	5	NC Teacher 1	3	5
Lagged: Cognitive	0.924	0.913	0.924	0.059	0.042	0.044
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Lagged: NC Teacher	0.001	0.007	0.004	0.441	0.489	0.526
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Nonlag: Investment	0.004	0.024	0.02	0.088	0.12	0.029
	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)
Nonlag: Background	0.029	0.041	0.032	0.001	0.028	0.035
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)

Table 18: Production function. Waves K-5. Standard errors in parentheses. Specification includes parental investment and parental background. Allowing for autocorrelated measurement errors. Investments are not allowed to be endogenous.

Grade	Cognitive 1	3	5	NC Teacher 1	3	5
Lagged: Cognitive	0.928	0.914	0.93	0.061	0.041	0.044
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Lagged: NC Teacher	0.003	0.007	0.006	0.44	0.489	0.527
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Nonlag: Investment	-0.013	0.017	-0.002	0.088	0.123	0.027
	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)
Nonlag: Background	0.042	0.047	0.047	-0.003	0.022	0.034
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)

Table 19: Production function. Waves K-5. Standard errors in parentheses. Specification includes parental investment and parental background. Allowing for autocorrelated measurement errors. Investments are allowed to be endogenous.

Grade	Cognitive 1	3	5	NC Teacher 1	3	5
Lagged: Cognitive	0.931	0.925	0.93	0.066	0.033	0.053
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Lagged: NC Teacher	0.003	0.01	0.01	0.435	0.476	0.509
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Nonlag: Investment	-0.037	0.046	-0.015	0.08	0.136	0.011
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
Nonlag: Background	0.033	0.054	0.048	0.002	0.011	0.05
	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)
Nonlag: Quality	0.043	-0.098	0.027	-0.023	-0.015	-0.035
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)

Table 20: Production function. Waves K-5. Standard errors in parentheses. Specification includes parental investment, parental background, and school quality. Allowing for autocorrelated measurement errors.

	NC Parent
Grade	1
Lagged: Cognitive	0.06
	(0.03)
Lagged: NC Parent	0.633
	(0.03)
Nonlag: Investment	0.101
	(0.04)
Nonlag: Background	-0.039
	(0.03)
Nonlag: Quality	0.047
	(0.04)

Table 21: Production function. Waves K-5. Standard errors in parentheses. Specification includes parental investment, parental background, and school quality. Allowing for autocorrelated measurement errors.

	black			hispanic			asian		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
K	-0.868	-0.868	0	-0.817	-0.817	0	-0.292	-0.292	0
	(0.042)	(0.042)	(0)	(0.064)	(0.064)	(0)	(0.128)	(0.128)	(0)
1	-0.905	-0.887	-0.08	-0.804	-0.787	-0.029	-0.328	-0.337	-0.066
	(0.048)	(0.042)	(0.033)	(0.062)	(0.062)	(0.028)	(0.118)	(0.109)	(0.042)
3	-1.007	-0.934	-0.196	-0.784	-0.713	-0.019	-0.205	-0.194	0.053
	(0.041)	(0.04)	(0.04)	(0.067)	(0.068)	(0.03)	(0.123)	(0.111)	(0.053)
5	-1.1	-0.998	-0.311	-0.718	-0.617	0.029	-0.111	-0.097	0.133
	(0.04)	(0.041)	(0.036)	(0.07)	(0.07)	(0.034)	(0.099)	(0.085)	(0.052)
Controls for investment		x	x		x	x		x	x
Deviation from Expectation			x			x			x

Table 22: Racial gaps for cognitive skills, raw values and counterfactuals. Waves K-5. Specification includes parental investment, parental background, and school quality. Allowing for autocorrelated measurement errors. Standard errors from bootstrap with $R = 10$ iterations.

	black			hispanic			asian		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
K	-0.31	-0.31	0	-0.068	-0.068	0	0.16	0.16	0
	(0.111)	(0.111)	(0)	(0.076)	(0.076)	(0)	(0.062)	(0.062)	(0)
1	-0.353	-0.245	-0.056	-0.08	0.023	0.102	0.264	0.347	0.295
	(0.087)	(0.071)	(0.041)	(0.067)	(0.075)	(0.071)	(0.082)	(0.09)	(0.08)
3	-0.523	-0.307	-0.181	-0.021	0.182	0.252	0.358	0.501	0.487
	(0.103)	(0.109)	(0.102)	(0.071)	(0.083)	(0.074)	(0.173)	(0.165)	(0.158)
5	-0.43	-0.261	-0.162	0.02	0.184	0.252	0.517	0.613	0.616
	(0.09)	(0.102)	(0.101)	(0.029)	(0.052)	(0.051)	(0.043)	(0.063)	(0.073)
Controls for investment		x	x		x	x		x	x
Deviation from Expectation			x			x			x

Table 23: Racial gaps for noncognitive skills, raw values and counterfactuals. Specification includes parental investment, parental background, and school quality. Allowing for autocorrelated measurement errors. Standard errors from bootstrap with $R = 10$ iterations.

Women	Raw		
	Black	Hispanic	Asian
25-29	-0.89** (0.026)	-1.13** (0.032)	1.083** (0.046)
30-34	-0.952** (0.028)	-1.279** (0.037)	1.396** (0.057)
35-39	-0.831** (0.026)	-1.187** (0.037)	1.146** (0.063)
40-44	-0.741** (0.024)	-1.158** (0.038)	1.018** (0.07)
Men	Raw		
	Black	Hispanic	Asian
25-29	-0.994** (0.026)	-1.099** (0.032)	1.154** (0.045)
30-34	-1.027** (0.028)	-1.146** (0.036)	1.362** (0.054)
35-39	-1.016** (0.027)	-1.21** (0.038)	1.259** (0.06)
40-44	-0.96** (0.025)	-1.155** (0.039)	1.223** (0.069)

Table 24: Racial Gaps in Years Of Education, controlling for year dummies. Source: ACS 2006-2010. Universe: US-born respondents.

Women	Raw			Via Edu		
	Black	Hispanic	Asian	Black	Hispanic	Asian
25-29	-0.157** (0.009)	-0.094** (0.012)	0.253** (0.016)	-0.085	-0.103	0.089
30-34	-0.204** (0.01)	-0.14** (0.013)	0.342** (0.019)	-0.108	-0.14	0.145
35-39	-0.154** (0.009)	-0.099** (0.013)	0.325** (0.022)	-0.087	-0.121	0.136
40-44	-0.14** (0.009)	-0.114** (0.014)	0.31** (0.024)	-0.076	-0.129	0.124
Men	Raw			Via Edu		
	Black	Hispanic	Asian	Black	Hispanic	Asian
25-29	-0.224** (0.009)	-0.124** (0.011)	0.151** (0.016)	-0.054	-0.082	0.095
30-34	-0.289** (0.01)	-0.16** (0.012)	0.225** (0.018)	-0.079	-0.115	0.15
35-39	-0.328** (0.01)	-0.208** (0.013)	0.249** (0.021)	-0.096	-0.128	0.148
40-44	-0.336** (0.009)	-0.253** (0.014)	0.23** (0.024)	-0.096	-0.126	0.148

Table 25: Racial Gaps in Log Hourly Wage, controlling for year dummies. Source: ACS 2006-2010. Universe: US-born respondents, not in school.

	Rate White	Rate Black	Rate Asian	Rel. Rate Black	Rel. Rate Asian
Total	3803.91	8413.61	1405.34	2.21	0.37
Murder and manslaughter	1.46	9.71	0.42	6.68	0.29
Forcible rape	5.75	14.55	1.80	2.53	0.31
Robbery	30.07	301.87	17.52	10.04	0.58
Aggravated assault	83.44	296.99	27.77	3.56	0.33
Burglary	138.13	392.85	34.25	2.84	0.25
Larceny-theft	630.65	1435.15	296.82	2.28	0.47
Motor vehicle theft	32.36	120.35	12.78	3.72	0.40
Arson	12.34	15.39	3.60	1.25	0.29
Violent crime	120.71	623.13	47.51	5.16	0.39

Table 26: Arrest rates below age 18 by race and crime in 2009. Data from FBI (`http://www2.fbi.gov/ucr/cius2009/data/table_43.html`). Rates are defined as arrests per 100,000 juveniles. Relative rates ("Rel. Rate") is the rate for a racial group divided by the rate for whites. The data do not allow analysis by gender, birthplace, or family background.

	Cognitive	Noncognitive
Education	1.303** (0.044)	0.216** (0.043)
Log Wage Age 30	0.119** (0.01)	0.03** (0.01)
Log Wage Age 34	0.142** (0.012)	0.055** (0.011)
Log Wage Age 38	0.148** (0.01)	0.043** (0.01)

Table 27: Effect of skills on economic outcomes. Data: BCS 1970. Skills measures defined to correspond to measures in ECLS-K. Cognitive: Average of math and reading test. Noncognitive: Average of teacher-ratings of interpersonal behavior, externalizing, internalizing, and self-control. All measures taken at age 10 (corresponding to the ELCS-K gaps at in grade 5,age 10-11, and normalized to Z-scores. Universe: Full sample for years of education. For log wage, restricted to respondents who work at least part-time. Controls for gender, parental background, and region of birth.

Gaps in Skills	Cognitive	Noncognitive		
black	-0.794	-0.23		
hispanic	-0.193	0.14		
asian	0.044	0.524		
Predicted Gaps in Outcomes	Education		Wage	
	Cognitive	Noncognitive	Cognitive	Noncognitive
black	-1.149	-0.055	-0.131	-0.011
hispanic	-0.279	0.034	-0.032	0.007
asian	0.063	0.126	0.007	0.025

Table 28: Racial gaps predicted from skill gaps. This table predicts the gaps in outcome using the production function estimated with BCS70 data. Skill gaps are estimated using the ECLS-K. All regressions control for family background.

B Appendix

B.1 Panel Structure and Attrition

The ECLS-K has a complicated sampling scheme. For instance, only a random sample of school movers was followed up in later surveys. In addition, numerous schools were dropped from the survey over time. Furthermore, attrition is high.

I proceed as follows: I drop all individuals who attrite at some time during the sample period, thus obtaining a balanced panel of 8722 individuals.

In this paper, I will not use sampling weights to correct for attrition on covariates. In earlier versions, I estimated all my models using sampling weights.[12] This requires calculating the covariance matrix of all measures with sampling weights, which was computationally tenuous. Nonetheless, the estimates using the weighted observations were similar to non-weighted observations, thus I omit the sampling here.

In addition, I have attempted to test whether the assumption of conditionally random attrition is tenable. To do so, I estimated a two-stage selection model: An outcome Y is related to covariates X via the equation

$$Y = \beta X + e$$

with e being independent of X. However, the econometrician's sample is selected because Y and X are only observed if the observation stays in the sample. Non-parametric identification of selection is

[12] In this paper, I will account for attrition by using the sampling weights provided in the data. This approach allows recovering the representative distribution of all variables, assuming that attrition is conditionally missing at random; where the conditioning controls for a range of features, such as location, school type, age and race (Tourangeau et al., 2009).

possible if the selection decision is modelled and instruments are available (Heckman and Vytlacil, 2007). I model attrition via the equation

$$D = I\left(\gamma_1 Y + \gamma_2 X + \gamma_3 Z + \varepsilon > 0\right)$$

where the instruments Z need to be independent of the residuals e and ε. Plausible instruments are characteristics of the interview, especially measures of the time (or psychic) cost of participating in the survey (Fitzgerald et al., 1998). Fortunately, such measures are available: I create two variables

- Number of re-contacts to complete the interview
- Total duration of interview

However, none of these predict attrition. Therefore, the selection (on unobservables) into attrition is not identified nonparametrically and I do not undertake further analysis of this issue.

B.2 Structural Equation Models

B.2.1 Allowing for Autocorrelated Measurement Error

Why do autocorrelated measurement errors matter?

For a simple illustration, assume that there is only one factor, two measurements, and two time periods. The factor loadings are unity for all measures, and the factor variance is .5. This can result in a covariance matrix as follows:

$$\begin{array}{c|cccc} & Y_1^1 & Y_1^2 & Y_2^1 & Y_1^2 \\ \hline Y_1^1 & 1 & & & \\ Y_1^2 & .5 & 1 & & \\ Y_2^1 & .25 & .25 & 1 & \\ Y_2^2 & .25 & .25 & .5 & 1 \end{array}$$

This matrix identifies the γ in

$$\theta_2 = \gamma\theta_1 + \eta_2$$

because

$$\gamma = \frac{Cov\left(Y_1^1, Y_2^2\right)}{Cov\left(Y_1^1, Y_1^2\right)} = \frac{Cov\left(Y_1^1, Y_2^1\right)}{Cov\left(Y_1^1, Y_1^2\right)}$$

in this case, this yields

$$\gamma = .5$$

However, assume that measurement error (for a given measure) is correlated over time. The covariance matrix now looks as follows:

$$\begin{array}{c|cccc} & Y_1^1 & Y_1^2 & Y_2^1 & Y_1^2 \\ \hline Y_1^1 & 1 & & & \\ Y_1^2 & .5 & 1 & & \\ Y_2^1 & .8 & .25 & 1 & \\ Y_2^2 & .25 & .8 & .5 & 1 \end{array}$$

and no longer does the matrix yield a unique estimate for γ, because:

$$\frac{Cov\left(Y_1^1, Y_2^2\right)}{Cov\left(Y_1^1, Y_1^2\right)} = .5 \neq 1.6 = \frac{Cov\left(Y_1^1, Y_2^1\right)}{Cov\left(Y_1^1, Y_1^2\right)}$$

One of the estimates is wrong, and overstates γ. If γ is overstated, this can cause the optimizer to assume that the data is best described by setting $V\left(\eta_2\right) = 0$.

For this reason, I allow for autocorrelated measurement error. I model that

$$\varepsilon_t^j = \rho^j \varepsilon_t^j + \widetilde{\varepsilon}_t^j$$

where $\widetilde{\varepsilon}_t^j$ is independent over time and between measures. In the appendix, I show that ρ^j is identified under weak assumptions: the main restriction is that one needs at least two measures for a factor in every time period.

Implementing the autocorrelated error is a small statistical change, but makes it infeasible to still use the Kalman filter (because the Kalman filter assumes that measurement error is uncorrelated over time). Instead, I estimate the model by describing it as a structural equation model. The statistical model is implemented using the *OpenMX* software package in *R*. For computability, I assume normality of all errors, but this is not needed to achieve identification.

With convergence assured, I am more confident that cross-effects are well-interpretable. Their magnitudes are modest: past investment increases skills, and past skills increase investment, but the partial correlation has an absolute value below .05 in both cases.

B.2.2 Estimation of the Structural Equation Model

The basic model consists of measures and latent factors. For expositional purposes, I simplify to assume that $T = 2$. Then the observations result from

$$\begin{aligned} Y_t &= \alpha_t \theta_t + \varepsilon_t \\ \theta_t &= G_t \theta_{t-1} + \eta_t \end{aligned}$$

I now re-write this model in the "RAM" (Reticular Action Model) notation, which directly maps into the likelihood. Define

$$v = \begin{pmatrix} Y_1 \\ Y_2 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

then one can write:

$$v = Av + \widetilde{\varepsilon}$$

with

$$A = \begin{pmatrix} 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & \alpha_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & G & 0 \end{pmatrix}, V(\widetilde{\varepsilon}) = \begin{pmatrix} V(\varepsilon_1) & 0 & 0 & 0 \\ 0 & V(\varepsilon_2) & 0 & 0 \\ 0 & 0 & V(\eta_1) & 0 \\ 0 & 0 & 0 & V(\eta_2) \end{pmatrix} = \begin{pmatrix} K_1 & 0 & 0 & 0 \\ 0 & K_2 & 0 & 0 \\ 0 & 0 & Q_1 & 0 \\ 0 & 0 & 0 & Q_1 \end{pmatrix}$$

The likelihood results as follows: one observes the first two components of A, I call these

$$\widetilde{v} = jv, j = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and their variance is given by

$$V\widetilde{v} = jVvj' = j(I-A)^{-1} V(\widetilde{\varepsilon}) \left((I-A)^{-1}\right)' j'$$

And because normality is assumed throughout, $\widetilde{v}$ is jointly normal, thus its likelihood is given by

$$f(\widetilde{v}|\text{mean} = 0, \text{V} = V\widetilde{v})$$

where f is the density of a normal distribution.

The extension with regard to intercorrelations of measurement errors or factors are easily incorporated in this notation. For instance, autocorrelated measurement errors introduce a new latent variable, $\widetilde{\varepsilon}_t$.