OTDK-dolgozat

2017

Kiss Olivér

Budapesti Corvinus Egyetem

Rózemberczki Benedek András

Budapesti Corvinus Egyetem

Fekete és fehér gyermekek teszt pontszámainak különbségeiről a No Child Left Behind Act után

The black and white score gap after the No Child Left Behind Act

Konzulens: Bozsér Kata

Kézirat lezárásának dátuma: 2016.03.23.

"This is not only an American problem. In the Turkish ghettos of Vienna, in the Somali ghettos of East London there are children in the very same situation."

(Roland Fryer - Calvo-Armengol Prize Lecture, 2010)

Abstract

The study investigates the black and white test score gap among elementary school students in the United States after 2010. A core finding of the research is that black students do worse than whites in mathematics even when environmental variables such as socio-economic status are controlled for. The size of the raw score gap in the fall of kindergarten in mathematics is 0.526 standard deviations, with the use of controls this shrinks to 0.158 standard deviations. The score gap widens considerably among black and white students in the first two years of the formal education. The raw mathematics score gap in the spring of elementary school is 0.635 standard deviations, after controlling for a handful of factors it is reduced to 0.308 standard deviations. We also found that after the use of propensity score matching the mathematics score gap in the fall of kindergarten becomes insignificant. Intriguingly, even after propensity score matching the gap starts to widen, which might flag the results of racial homophily or consequences of premarket discrimination of black students.

JEL Classification: I24, J15, J16

Keywords: Racial achievement gap, Test scores, No Child Left Behind Act

Contents

List of Figures

List of Tables

Chapter 1: Introduction

The pre-market achievement gap among black and white students in the United States has been in the focus of many studies since Coleman et al. (1966). It is a well known empirical regularity that standardized score results of black and white elementary school students start to diverge early on. As the work of Fryer and Levitt (2004) emphasized it, the divergence is the most severe in case of mathematics. The difference in the students' achievements manifest as early as kindergarten even in case of the most basic numerical skills such as counting, adding and subtracting numbers. These very basic skills serve as the foundation for other, more advanced skills such as calculus, linear algebra and probability that are very much needed for obtaining an advanced degree. As black students have a setback early on in these skills, the majority of them is later excluded from the quantitative niche labor markets such as finance, engineering and computer science (Franklin, 1999). The authors of this paper believe that this phenomenon should not be overlooked, because the premarket trajectory of the achievement gap is an indicator of the future labor market achievement gap which will be present in the United States.

As Fusarelli (2004) points out, the regularity outlined in the previous paragraph had a large impact on the education policy of the United States. The reduction of the black and white test score gap was a clear goal of the *No Child Left Behind Act*, which was enacted in 2001 (henceforth abbreviated as NCLB). The positive impact of the NCLB is rather ambiguous according to a number of authors, including Darling-Hammond (2007), Dee and Jacob (2011) and Gaddis and Lauen (2014). The comprehensive investigation of Fryer and Levitt (2004) measured the test score gap between 1998 and 2000, just a few years before the act was authorized. The dataset that we used for the research is from the exact same source and was compiled with identical methods, but it has data about children who entered kindergarten in 2010. This means that it allows for the quantification of relative gains that black students made until the end of the last decade. The setting is not optimal (it is far from the evaluation of a randomized ceteris paribus policy intervention), but an overall assessment about the change in the black and white test score gap can be performed.

The benchmarks of our later comparison are the quantitative results of Fryer and Levitt (2004). They found that the raw normalized mathematics score gap between black and white students was -0.638 standard deviations, while the raw reading score gap was -0.401 standard deviations among preschoolers in 1998. When family inputs and school quality were controlled, the gap narrowed, but it did not disappear. Intriguingly they also found that both the raw and conditional score gaps significantly widened among black and white students by the end of elementary school's first year. This paper is in line with the mainstream of the previous literature on the black and white score gap, which we define as work related to the debate started by Coleman et al. (1966). The paper at hand investigates the following, somewhat closely related questions:

1. *How large was the achievement gap among black and white kindergarten and elementary school students who started preschool in 2010/2011?*
2. *What are the possible causes of the achievement gap?*
3. *Under which conditions does the gap disappear?*
4. *Has the black and white test score gap become smaller or greater since 1998?*
5. *Was the No Child Left Behind effective in terms of closing the relative gap?*

The remainder of this paper is structured as follows: a brief overview of the literature about the black and white score gap and policy interventions is given in Section 1.1. The dataset used for the analysis and the data cleaning-handling process is presented in Section 1.2. of this chapter. The introduction of the dataset is followed by a brief presentation of descriptive statistics in Section 1.3. The econometric methods used in the paper are discussed in Chapter 2. The results of the regression models are presented in Chapter 3, and the findings of more advanced methods can be found in Chapter 4. The concluding remarks are made in Chapter 5. The list of notations is in Appendix A, longer tables can be found in Appendix B. The authors of the paper support the norms of reproducible research, therefore the cleaned dataset and STATA codes are available at `https://rozemberczki.wordpress.com/`.

1.1 The context of the score gap and interventions

Investigations about the black and white score gap sometimes have implicit assumptions, that are not made explicit by the authors. In addition, a number of specifications used for measurement have certain weaknesses. These assumptions and common problems with the identification strategies are summarized by Quinn (2015a) and Quinn (2015b). These are the following:

1. When authors investigate the dynamics of the gap – increases and decreases through time – they have to use a standard deviation measure. This measure used for normalizing the changes effects the significance of the parameters as Quinn (2015a) highlights it. Therefore it is clear, that the first differences divisor used for normalization matters – luckily our approach is different, because it is not dynamic. This means that we do not have to deal with this problem.

2. The scale nature of the tests is pinpointed by Quinn (2015a) – as he argues, the linear data generator process (the fact that return on family and school inputs is constant) is a latent assumption in most of the models. We explicitly state that this is an assumption in our methodology – the marginal improvements are the same for every unit of a certain input.

3. The test date independence assumption – the proper measurement of the black and white score gap would assume that the the testing date is not correlated with score racial composition of school. In plain words, authors should assume and test whether the black majority schools make tests earlier. Throughout our analysis we assume that the test dates and the racial composition of schools and kindergartens are independent.

4. Even in cross section models (which is our approach) the significance of estimates depend on the estimation of standard errors. Previous papers in the literature used plain standard errors or robust errors. As students' achievement within a school is likely to be correlated, we propose the use of school clustered standard errors. We argue for this choice later in Chapter 2.

It is worth emphasizing that the black and white score gap is quite robust regarding the introduction of controls. The initial results of Fryer and Levitt (2004) had shown that the conditional gap is significant if one uses school fixed effects and family input controls. Their later results had shown that the conditional gap is becoming wider in higher grades of elementary school (approximately starting at 3^{rd} grade). This type of trajectory is quite typical in case of other countries – an example can be the Roma in Hungary. The Roma and non-Roma conditional score gap in reading and in mathematics is known to widen with years (Kertesi and

Kezdi, 2011). However, as Kertesi and Kezdi (2015) had shown, even the later year gaps are insignificant when matched regressions with controls are used. Our results will show that this is not the case when the black and white score is investigated in years starting from 2010. The divergence is still significant in the propensity score matched linear regressions.

The initial black and white score gap results in a long-term and widening ability and achievement gap among black and white adults. Based on the results of Fryer (2011) we know that the black and white wage gap is approximately 28% among black and white males. Importantly, this relative difference in wages was reduced to 0.6% when the standardized test results in mathematics were used as controls. This shows that the differences in test results have persistent effects on the later labor market outcomes of black students. It also underpins our previous argument that the mathematics score gap deserves exceptional consideration. The reading scores are also investigated by Fryer (2011). One of the most shocking facts about the reading scores is that in Chicago only 3% of black 8^{th} graders can read at grade level and this ratio is only 1% for the 12^{th} grader black students. This has troublesome implications about later life outcomes.

As Fryer and Dobbie (2011) point out a number of micro-level measures used as control in the investigation of the black and white score gap is ill-suited. Variables such as the student to teacher ratio or lumpsum investments to student ratios are ineffective measures of school quality. These measures do not flag that the teachers and students in a given school are incentivized to perform better. About the weak measurement value of teacher to student ratio Fryer and Dobbie (2011) give an intuitive explanation. If a school has a bad reputation, it would have an extremely low number of pupils and this would result in a relatively good student to teacher ratio. The existence of such outliers would imply that the school performance is uncorrelated with the student to teacher ratio. In our investigation of the black and white score gap, to hedge this problem, we mainly use micro level input variables from the family's and school's side.

Our paper partially focuses on the effectiveness of the NCLB program. The NCLB is in part an incentivization and accountability based education intervention program (Gaddis and Lauen, 2014). These programs rely on the incentivization of both students and teachers. The possible incentivizations of the students and teachers in an education intervention program in most cases fall into one of the following two categories:

1. **Output based incentives:** This type of incentivization connects rewards to the final outcome. From a regulation point of view, this type of incentivization is perfect when the

producers know the production function and they also know the optimal input combination. In our case, the output based incentivization means that students are rewarded for high achievements (but only in the end) and teachers are rewarded when their classes are doing well. In addition, we would assume that students know what is the best learning technique (e.g. flashcards, solving exercises or reading theory) and optimal amount spent with their studies. Essentially the NCLB used output based incentives – schools that were not doing well were closed and high performing teachers were rewarded.

2. **Input based incentives:** This type uses a different mechanism. Rewards are given after micro-level inputs that are done by the teachers and students for achieving the objective (passing the exam with a high score or obtaining a given set of skills). This incentivization gives teachers credit for tutoring and teaching extra classes. It also gives students micro-level incentives (money on a safe-account, food, clothing) if children solve a set of math tasks, get through with readings, attend school and do not disrupt the classes. As (Bradley and Fryer, 2011) puts it, it is an optimal policy when the students (producers) do not know the steps that lead to a high score on a mathematics test.

Importantly, the higher effectiveness of input based incentives were proved by the randomized experiments that Bradley and Fryer (2011) and Fryer (2013) concluded. The effectiveness of these policies is large both in case of teachers and students. We can state that the external validity of these randomized experiments would be strong for the United States[1]. It is worth emphasizing that the NCLB was a program that relied on input based incentives, which might have constrained its effectiveness as we will see later.

1.2 The dataset

The analysis is based on the Early Childhood Longitudinal Study Kindergarten Cohort 2010/2011 (henceforth ECLS-K 2010/2011) panel dataset, which was released by the *National Center for Statistics* in April 2015. This dataset is representative regarding the United States regarding racial composition of students, and connects information about the school performance (test results) and family background of children who entered kindergarten in the school year 2010/2011. It also includes variables that describe the quality of the kindergarten and school attended by the children. Essentially, the survey is a linked student-school dataset[2]. Altogether 18,174 observations are included in the survey, as all the participants were to be tested in the fall

[1]The authors assume that the external validity would be also high outside the United States.

[2]LSSD

and spring of kindergarten, as well as at the end of the first school year. Similarly to the ECLS-K 1998/1999 dataset, a random subsample of students is tested in the fall of the first school year.

The original dataset is a wide panel, and the number of potential control variables is well beyond 5,000. To reduce the dimensionality of the dataset we arbitrarily picked potential predictors based on previous research and our intuition. The list of extracted variables used for the analysis is attached in Appendix B as Table B.1. In the whole sample, 1,699 students have missing racial attribute values, while 12,402 students are white and 2,792 students are black. There is additional information about being Asian or Hispanic, and the *Race else* variable describes the pooled group of Pacific Asians and Native Americans. An other time invariant characteristic included is the child's gender. Later this variables is included as control on the right hand side.

The achievement of students is measured on a scale of 0 to 80, both in reading and mathematics. The two scores came from 4 exercises (e.g. in reading, "Simon says" is one of the tasks) that account for 20 points in the overall score. These variables will be on the left hand side as dependents. The psychometry of the tests is designed in a way that the return on skills is non-linear – as Quinn (2015a) points out this results in the phenomenon that the distribution of the test scores is symmetric. The attrition regarding these variables is quite heavy: 2,505 students were not tested in mathematics in the fall of kindergarten, while in the spring of the first elementary school year this number was 3,059. Similarly, 2,579 students were not tested regarding reading skills in the fall of kindergarten, while in the spring of the first elementary school year this number was 3,071. The attrition would seriously weaken the external validity of possible later findings, therefore students with missing score data were excluded from our sample. This selection was done in the most stringent way. For example, if a student had a missing score in the fall term of 2010 for reading, she or he was dropped from the sample used for estimating the regression on mathematics in the different terms.

General socio-economic background (henceforth SES) of the child's family is described by a composite measure that includes the parents' profession and the household income. This variable is a pre-normalized continuous measure, which is symmetric. As a result of this, log transformation is not needed, unlike other income measures would require, as they regularly show strong skewness, a tail on the right. This variable was constructed by the data providers to have a zero mean and unit standard deviation. It has to be emphasized, that this is the only continuous variable that is remeasured at the observation points.

The child's weight at birth (measured in ounces) and the family's participation in the *Program for women, Infants and Children* (hereafter abbreviated as WIC) are important possible poverty indicators that might influence later achievements. The age in months when a child enters kindergarten or school is implicitly influential, because of this, it is included as a control variable. Another advantage can be the number of children's books owned by the family – this is also extracted. It has to be noted that these control variables were also used by Fryer and Levitt (2004, 2006) and Gaddis and Lauen (2014) when the racial gap based on ECLS-K 1998/1999 was investigated. The mother's age at the birth of the first child in the family might be also influential. However, we are not sure whether the change in expected achievement associated with the mother's age is linear. Because of this, we defined two dummy variables according to:

1. If the mother was under 18 at the birth of first child, the variable young mother is 1, otherwise it is 0.

2. If the mother was above 30 at the birth of the first child, the variable mature mother is 1, otherwise it takes the value 0.

The school quality might also have an impact on the achievement of students, and because of this, we extracted a handful of potential control variables that describe the quality of a school. We generated dummy variables for each of the individual schools – this allows for taking out school fixed effects. To measure the general poverty level of the student body we included the variable free lunch, which measure the percentage of students receiving a state funded lunch. The variable public school describes whether the school is a private institution or a public one. If the school is state owned, this variable takes the value of one. The racial composition of the school is measured by the ratio of black and white students. To avoid collinearity, we excluded the ratios of students from other races. The ordinal variable turnover describes the turnover of the teachers on a scale of 0 to 3, where the lowest value means that there is no problem with the frequent turnover of teachers. The security dummy describes whether the school security uses a metal detection gate or not. The internet access variable is also a dummy, which is zero when the school has no internet access. The computers variable describes the number of computers per student in a continuous way. The ordinal variables drug and gang problems describe whether the school struggles with a drug or gang problem. The above mentioned ordinal variables are all used solely for generating dummy variables. In the modeling phase, only the dummies were included.

The work of Kertesi and Kezdi (2015) had shown that early age parental inputs such as story telling have a strong impact on expected school achievements of minority students. Leaving out such predictor variables would lead to endogeneity due to omitted variable bias (Wooldridge, 2010). As we want to avoid omitted variables, we gathered the following controls that might describe parental inputs to a certain extent. The reading time variable describes the average time spent by the parents and children reading together. It is a continuous measure in the survey, but parents rather report integer values such as 20 minutes and so on. The telling stories and book reading variables are ordinal, and describe the frequency of the mentioned activities – higher values represent a higher frequency of the activity. The computer home variable describes whether a family has a computer in their home that is accessible to the child. The ordinal variables are transformed into dummies in the later analysis.

1.3 Descriptives of data

A brief overview of the observed empirical regularities in the data might help the later econometric modeling. The race specific mean scores[3] on level are included in Appendix B as Table B.3. This table shows that when students enter kindergarten, black students are behind Asian and white students in mathematics, but they somewhat do better than Hispanic students. The same stands for the reading scores. Importantly, at the end of the first school year the Hispanic students do better than blacks. The leading position of white and Asian students is unchanged, and the relative gaps among black, white and Asian students are also unchanged. This holds until the next year based on these simple descriptive statistics. It is worth mentioning that the race specific test scores show somewhat different distributions. On Figure 1.1, Gaussian kernels of the white and black mathematics test scores are plotted. The distribution of the white test score shows bimodality, and has a tail on the right, while the distribution of black students' score is rather symmetric. However, the different group specific mean scores are quite prevalent. Importantly, the scores on the level are hard to interpret. To resolve this later on, we use normalized test scores which allows for cross-time comparison. This also makes possible the better evaluation of relative racial gains and losses.

Another aspect that should not be overlooked is that students of different race come from families and schools that have fundamentally different qualities. In Table **??** of Appendix B, the mean and standard error of continuous variables is calculated for the races separately. It turns out that in the sample, parents of Asian students have the most prestigious occupations which pay the highest salaries. The lowest paying jobs are held by the parents of Hispanic students, and the average of black families is also nearly 0.470 standard deviations below the mean.

[3] Here the parentheses contain standard deviations – not standard errors.

Intriguingly, the socio-economic status of non-white families worsens in the observed time periods. The white families have the largest number of children's books, they own roughly 96 of them. The black families have the lowest number of children's books, the average black family has 45 of such books, which seems to be a rather serious constraint regarding early childhood development. A fascinating phenomenon is that Asian families are also significantly behind white families regarding books. The black mothers are the youngest in the sample, but they are not significantly younger than Hispanic ones. In addition, black children have the lowest average birth weight, they are even behind Asian students.

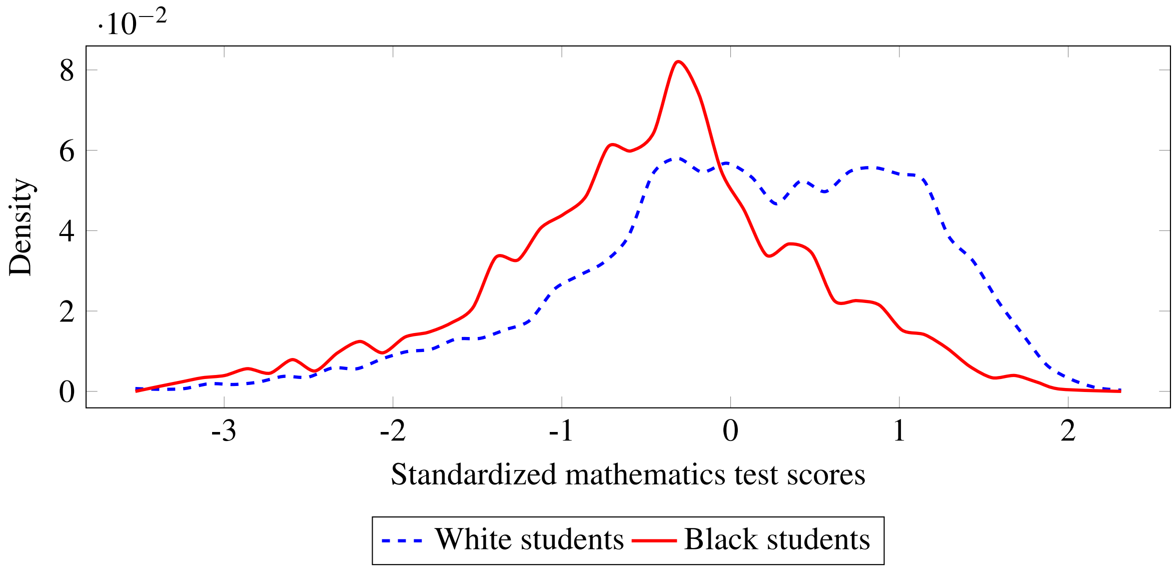

Figure 1.1: Distribution of mathematics scores in the spring of 2011

The school quality behind black students is also weaker to a certain extent. The average black student goes to an elementary school where approximately 55% of students is also black. A similarly assortative or selective school choice can be observed in case of white students. The average white student is in a school where roughly 63% of students is also white. The ratio of black students is well below 10% in the schools that the white average student's parents choose for her or him. This also implies that on a national average white students attend white majority schools, while black students attend black majority schools. This might either be a sign of segregation, assortative school choice or a mixture of the two. This is a rather strange finding, because the NCLB was explicitly against segregated schooling of black students. As the average black student's socio-economic background is lower, a larger number of these students is eligible for a state funded free lunch. This is clearly shown by the grouped average ratio of student's receiving free lunch.

Chapter 2: Methods

This chapter gives an in-depth overview of the statistical techniques that we use throughout the study for measurement and program evaluation. Multiple chained imputation, the semi-supervised machine learning method used for filling the missing values is the topic of Section 2.1. The estimated linear regression models are presented in Section 2.2. The interpretation of the parameters is discussed in case of every estimated equation. More advanced econometric techniques such as decomposition methods and propensity score matching are dealt with in Sections 4.2 and 2.4.

2.1 Multiple chained imputation

The variables in the dataset have multiple missing values and due to this, the panel is slightly unbalanced. As our main interest is in the relationship of achievement and race, we drop those observations that have missing racial attributes or achievement testing values. In addition, observations with missing gender values were also dropped from the sample. After these data handling steps a large number of observations still have missing values. Running regression on the unbalanced dataset would undermine the reliability and our findings would be less comparable across years, because there could be an endogenous reason behind the missing values. The variables that have missing values can be categorized as:

- *Binary variables*, the use of a metal detection security gate is a typical binary variable which was sparsely labeled in the dataset.
- *Ordinal variables*, for example certain parents did not report the frequency of telling stories and reading books in the family.
- *Count variables*, the number of books in the households has missing values, and this variable is a typical count variable, because the number of books is always integer.
- *Continuous variables*, in multiple cases the socio-economic status of the family was missing in the later years.

The missing value problem can be solved with multiple imputation by chained equations (henceforth MICE), which is from a point of view an ensemble semi-supervised learning method. As Hastie et al. (2009) point out, regression models have low bias and high variance, which means that an ensemble of the weak regression models leads to a lower predictive error. This is the reason why we chose MICE over the simple imputation. Based on Azur et al. (2011), we implemented the most general method of multiple imputation with chained equations that can be summarized as follows:

1. A variable with missing values is chosen to be on the left hand side of the imputation equation.

2. A proper regression equation is specified, with all the other variables from the same year on the right hand side. Based on the data type, the following models have to be chosen:

   (a) *Binary variables* – probit

   (b) *Ordinal variables* – ordinal probit

   (c) *Count variables* – Poisson regression

   (d) *Continuous variables* – linear regression

3. A regression model is estimated from the dataset with maximum likelihood or ordinary least squares.

4. The missing values are estimated with noise.

5. An other variable with missing values is chosen to be on the left hand side and the already imputed values are used for the inference on the missing values in the other variable.

6. The iterative process stops when no missing values can be predicted. Because gender, race and the achievement values have no missing values, the process converges to a balanced panel dataset in every case.

7. The number of iterative runs was 100, for each missing value we have 100 predictions with added noise. The missing values in the end were imputed according to the following rules of thumb:

   (a) *Binary and ordinary variables* – majority voting based on the mode of the imputed values.

   (b) *Count variables* – rounding the predictions mean over the imputations.

   (c) *Continuous variables* – taking the mean over the imputed values.

2.2 Regression models

Our primary interest is the expected achievement gap among black and white students. The dataset would allow for the use of panel data methods, but the measurement points are distributed unevenly in time. Because of this, we only investigate time specific linear regressions. To measure this, one can simply specify a linear regression for a given period which has the test scores on the left hand side and racial dummy variables on the right hand side. This naive model is described by Equation (2.1). The chosen reference group consists of white students, so the β_0 parameter shows their expected achievement while the other parameters show the expected difference among white and other student groups. We will focus on the β_1 parameter throughout the whole analysis. It's negative sign would show that black students' expected performance is behind the performance of white ones. The idiosyncratic error term is simply described by ε.

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \varepsilon_i \tag{2.1}$$

As the descriptive statistics had shown in Section 1.3, race is associated with certain variables that might affect later achievements such as birth weight , socio-economic status, the mother's age and participation in poverty programs. Omitting these variables from the estimated equation would result in biased parameter estimates. So we augment Equation (2.1) with a handful of controls such as: socio-economic status, number of children's books, gender, age of the child, weight, young mother dummy and participation in the WIC. This model is equivalent with the one that Fryer and Levitt (2004) estimated, so our results can be compared to theirs. The model is formally described by Equation (2.2), where γ denotes the vector of parameters and X denotes the controls.

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \text{X}_t'\gamma_t + \varepsilon_i \tag{2.2}$$

The β_1 parameter can be interpreted as follows: if two children have the exact same general background and they only differ in color (one of them is black, the other one is white), the expected achievement gap between them will be β_1. This equation might be augmented with school fixed effects (dummies that describe the schools). There is a simple rationale behind this idea. The school's quality might differ, and taking out the average school effect might account for a part of the achievement gap. This augmentation of Equation (2.2) is described in Equation (2.3), which is a specification used by Fryer and Levitt (2004). The parameter vector Θ, has a length that equals the number of schools in our sample.

$$\begin{aligned}\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \dots \\ + \text{X}_i'\gamma + \text{School}_i'\Theta + \varepsilon_i\end{aligned} \tag{2.3}$$

With the addition of school fixed effects, the interpretation of β_1 became the following: if two children have the exact same general background, they are in the same school and they only differ in color (one of them is black, the other one is white), the expected achievement gap between them will be β_1. Another approach to augment (2.2) is the addition of school input variables such as proportion of students receiving free lunch, ratio of black and white students, per capita number of computers, size of the classes, turnover of the teaching body, security measures of the school or categorical variables describing drug and gang related problems. In Equation (2.4) these variables are contained in Y, and the respective parameters are in the parameter vector Υ.

$$\begin{aligned} \text{Test Score}_i = &\beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \ldots \\ &+ \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \varepsilon_i \end{aligned} \tag{2.4}$$

As school inputs are included in Equation (2.4), the interpretation of β_1 becomes more nuanced: if two children have the exact same general background, the same school inputs and they only differ in color (one of them is black, the other one is white), the expected achievement gap between them will be β_1. The family inputs are possibly also influential, this implies that the inclusion of family specific factors reduces the parameter bias that is due to omitted variable bias. The specification defined by Equation 2.4 is further augmented by the family inputs in Equation (2.5). The family inputs are denoted by Z, and the respective vector of parameters is simply noted by Γ.

$$\begin{aligned} \text{Test Score}_i = &\beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \ldots \\ &+ \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \text{Z}_i'\Gamma + \varepsilon_i \end{aligned} \tag{2.5}$$

It is straightforward that the interpretation of β_1, the key parameter in our investigation changes accordingly: if two children have the exact same general background, the same school and family inputs and they only differ in color (one of them is black, the other one is white), the expected achievement gap between them will be β_1. The last augmentative step is the inclusion of the previously used school dummy variables that wash out the average time invariant school effect from the achievement of students. This is shown by (2.6), where the school fixed effects are also included. The interpretation of this fully specified model is quite complex: if two children have the exact same general background, they were in the same school, received the same school and family inputs and they only differ in color (one of them is black, the other one is white), the expected achievement gap between them will be β_1. This means that a large set of environmental factors is controlled fully in this specification.

$$\begin{aligned} \text{Test Score}_i = &\beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \ldots \\ &+ \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \text{Z}_i'\Gamma + \text{School}_i'\Theta + \varepsilon_i \end{aligned} \tag{2.6}$$

In the remainder of our study the standard errors of the parameter estimates are obtained by clustered errors, because simple standard errors or robust errors would be biased – for analytical details see Wooldridge (2010). It is a well founded assumption that the idiosyncratic errors are likely to be correlated in the same school class or kindergarten group. For example, let us imagine that in a school where only a single class was sampled, a teacher has a large positive impact on the achievement of the students. In this case the model would under or overestimate the score of students, and all of the residual errors would be positively correlated. This assumption also holds if classes consist of students who have similar skills or the assignment to elementary school classes is assortative. In addition, as the residual errors are positively correlated, the simple standard error or robust standard error estimation would underestimate the size of the standard errors. This would lead to false hypothesis testing, parameters that are small in absolute terms would appear to be significant. The clustering of the standard errors is done according to the school identification variable and we used the baseline implementation in STATA Cameron and Miller (2015). The value of the estimated parameters and the clustered standard errors are presented in the tables with three decimal digits.

2.3 Threefold achievement gap decomposition

As the descriptive statistics in Section 1.3 had shown, the initial family inputs of black and white students differ significantly. This fact implies that a fraction of the achievement gap might be caused by the difference in these inputs. To show this, in this section we introduce the counterfactual decomposition used by Blinder (1973) and Oaxaca (1973). This technique gives an upper estimate of the achievement gap that is due to possible premarket discrimination of black students. The derivation of the threefold decomposition presented in this section uses elements from the derivation of Andrisani and Daymont (1984). The method can be generalized to any mutually dichotomous pair of groups (whites and Hispanic students, males and females among many others).

The race specific linear regression models defined in Section 2.2 find a direct link between the expected value of the test scores and the predictors. One can simply define group specific regressions – respectively for the black and white students. With a general notation of the parameters (Ω) and the predictor variables, one can simply write these regression equations such as:

$$E(\text{Test score}_W) = E(\text{Controls}_W)'\Omega_W + \varepsilon_W \tag{2.7}$$

$$E(\text{Test score}_B) = E(\text{Controls}_B)'\Omega_B + \varepsilon_B \tag{2.8}$$

Based on the descriptive statistics presented in Section 1.3, it is known that the black and white students perform differently in mathematics and reading. The expected difference in performance of black and white pupils can be expressed as the difference of the group specific expected score values. This is simply the following expression for a given period and testing subject:

$$\Delta E(\text{Test score}) = E(\text{Controls}_W) - E(\text{Controls}_B) \tag{2.9}$$

The vector of parameters (Ω) and the control variables can be noted generally. The difference in expected scores can be expressed as a difference of the control variables at the mean. Based on Equations (2.7), (2.8) and (2.9), it comes that the difference is: l

$$\Delta E(\text{Test score}) = E(\text{Controls}_W)'\Omega_W - E(\text{Controls}_B)'\Omega_B \tag{2.10}$$

The expression defined by Equation (2.10) only holds exactly for the real (theoretical – non-estimated) random variables – the large sample estimates obtained from ECLS-K are paramount approximations of the expected difference in mathematics and reading scores. Finally, if the counterfactual expected values $E(\text{Predictors}_B)'\Omega_W$ and $E(\text{Predictors}_W)'\Omega_B$ are added and also subtracted from Equation (2.10), it is obtained that:

$$\begin{aligned}\Delta E(\text{Test score}) = {} & \underbrace{[E(\text{Controls}_W) - E(\text{Controls}_B)]'\,\Omega_B}_{\text{Difference in controls}} + \underbrace{E(\text{Controls}_B)'(\Omega_W - \Omega_B)}_{\text{Difference in sensitivity}} \\ & + \underbrace{[E(\text{Controls}_W) - E(\text{Controls}_B)]'\,(\Omega_W - \Omega_B)}_{\text{Difference caused by interaction}}\end{aligned} \tag{2.11}$$

The expression defined by (2.11) is the threefold Blinder-Oaxaca decomposition according to being black or white. The three counterfactual terms that are underbraced have substantive meanings:

1. **Difference in endowments:** The family background of black and white students, such as socio-economic status, differs strongly. A part of the achievement gap is caused by this difference in the family inputs received by the children – on average, black children's families are poorer and their environment induces weaker cognitive development than the environment of white students.

2. **Difference in sensitivity:** The black and white students respond differently to family inputs – their sensitivity to the controls variables is different. This expression contains also a part of the unobserved characteristics of the black students (e.g. parents love for the children) – this is only a possible upper estimate of the difference caused by the sensitivity due to omitted variable bias.

3. **Difference caused by the interaction:** There is an interaction between the difference in sensitivity and the difference in controls. This expression also contains the remaining part of the unobserved heterogeneity among the children.

It is clear that the 2. and 3. terms are more intriguing – these are the parts of the difference in the dependent variables that are induced by the plausible premarket discrimination of black students. However, these counterfactual terms only give an upper-estimate of possible discrimination – they give a bounding of the effect size regarding the controls. Importantly, this implies the following about premarket discrimination: if the 2. and 3. terms are insignificant, discrimination of black students is not present in the sample – observed differences in test scores are only caused by the differences in endowments. Another essential notion is that exclusive controls cannot be included in the model decomposition. Such variable can be a school fixed effect, when a school has only black or white students. The decomposition is only proper if Hispanic, Asian and students of other races are excluded from the sample. Otherwise, the decomposition would not show the proper counterfactual decomposition of the score gap among black and white students.

2.4 Propensity score matching

In an ideal evaluation study, the treatment (being black) is assigned randomly to the population – this defines an experiment that is randomized Imbens and Wooldridge (2009). This would ensure that the treatment is independent from other factors such as socio-economic status, school quality or birth weight among others. It is univoque, that a randomized experiment in this specific case is impossible. However, by applying propensity score matching such an ideal experimental setting can be imitated in a simple way Angrist and Pischke (2009). The propensity score matching method that we apply in order to quantify the *being black* effect on the achievement of the students can be summarized as:

1. The Hispanic, Asian and students of other races are excluded from the sample. The matching algorithm might match black students to other non-white minorities, because they have similar socio-economic background and that would lead to biased estimates.

2. We estimated a probit regression model in the following form: $\Pr(\text{Black} = 1 \mid W) = \Phi(W^T \eta)$. The dependent variable is the being black dummy, the predictors denoted by W were chosen as the general control variables, the family and school input variables (the set of controls that was used for the estimation of Equation (2.5).

3. A probabilistic prediction is done about being black, each student receives a propensity score of being black. For the predictions, we used the previously estimated model.

4. The black students each receive a white pair who has a similar propensity score based on simple nearest neighbor matching – essentially true positive blacks receive a false positive black pair (a white student, who has a similar probability to be black based on their endowments). We allowed for non-exclusive matching – the same white student might be assigned to be the pair of multiple black students.

We estimate a weighted version of the linear regression defined by Equation (2.12), where the weights were given by the propensity score matching process. Black students receive a weight of 1, white students receive a weight based on the number of matched pairs that they have. If a white student is matched up with none of the black students, she or he is excluded from the regression. If a white student is matched up with 5 black students, her or his weight is 5 in the estimation of the weighted linear regression.

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \text{Z}_i'\Gamma + \varepsilon_i \tag{2.12}$$

The β_1 parameter in Equation (2.12) gives a good estimate of the *being black* average treatment effect in different periods. If the probability of being black was positively correlated with the controls, the estimated β_1 parameter would be lower in absolute terms than in the linear regressions without matching.

Chapter 3: Measuring the score gap

This chapter presents the basic results of the research – simple linear regressions. The unconditional raw score gaps are analyzed in Section 3.1, while the score gaps with general socio-economic background controls are analyzed in Section 3.2, where estimates are also compared to the raw score gap. The family inputs are augmented with school fixed effects in Section 3.3. The inclusion of school inputs is discussed in Section 3.4, the model is augmented with family inputs in Section 3.5. A synthesis of the modeling approaches is carried out in Section 3.6, where the general controls, family and school inputs and school fixed effects are added to the estimated equation at the same time. The robustness of findings is thoroughly tested in Section 3.7. Comparison with the results of Fryer and Levitt (2004, 2006) is done in Section 3.8

The dependent variable *Test score* describes the standardized test score achieved by the students in a given period in mathematics and reading. The achievement gap is investigated in three periods – the fall of 2011, the spring of 2011 and the spring of 2012. As it was emphasized previously in Chapter 2, the time window among the measurement points is uneven. This would make the use and results of panel data methods questionable. Because of this, for the different time periods, simply different cross-section models are estimated. Intriguingly, the balanced nature of the dataset makes cross-time comparison possible, because the measurement is on the exact same population. We compare the results of the same cohort at different points in time. The remainder of the chapter focuses on the unconditional and conditional score gap among the students across time.

3.1 The unconditional test score gap

The unconditional racial score gap is measured by the linear regression that is defined by Equation (3.1). The dependent variable *Test score* describes the standardized test score achieved by the students in the given period in mathematics/reading – altogether we estimated 3×2 regressions. The standardization allows for cross-time comparison of the achievement gap. There are 5 dummy variables describing the race of the student – the white racial attribute is chosen as a reference category. The slope coefficients show the racial score differences. The parameter β_1 gives an unconditional estimate of the raw black and white score gap. In addition, while it is not of primary interest, the test score gap is obtained also for Hispanics, Asians and the pooled group of Native Americans and Pacific Asians.

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \varepsilon_i \tag{3.1}$$

The estimation results are included as Table 3.1, with robust standard errors in parentheses. The expected unconditional mathematics score gap between black and white students was -0.526 standard deviations in the fall of kindergarten. The reading gap in the same period was -0.290 standard deviations. The Hispanics and the pooled group performed worse than whites, while Asians performed better than whites in both subjects. Dismally, the achievement gap widens to the spring of kindergarten.

Table 3.1: Regression results – raw test score gap

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	0.198***	0.183***	0.229***	0.129***	0.134***	0.166***
	(0.013)	(0.013)	(0.012)	(0.014)	(0.013)	(0.012)
Black	-0.526***	-0.574***	-0.635***	-0.290***	-0.322***	-0.373***
	(0.026)	(0.027)	(0.026)	(0.026)	(0.026)	(0.027)
Hispanic	-0.644***	-0.554***	-0.598***	-0.486***	-0.468***	-0.519***
	(0.023)	(0.023)	(0.023)	(0.022)	(0.022)	(0.024)
Asian	0.281***	0.147***	0.094**	0.406***	0.259***	0.160***
	(0.039)	(0.036)	(0.033)	(0.048)	(0.041)	(0.034)
Race else	-0.473***	-0.329***	-0.407***	-0.439***	-0.370***	-0.344***
	(0.078)	(0.078)	(0.083)	(0.066)	(0.069)	(0.086)
N	13592	13592	13592	13592	13989	13592
R^2	0.093	0.082	0.097	0.055	0.053	0.060

Notes: Columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The racial attribute variables are used as regressors and multiracial students are excluded from the sample. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The fallback of black children in mathematics is -0.574 standard deviations, but the parameter is not significantly higher than the one obtained for the fall of kindergarten[1]. A similar drop behind is observed regarding the reading scores, to the spring of kindergarten the fallback is -0.322 standard deviations – this is also insignificant. The fallback and advantage of the other non-white groups is eased by the first spring. In the spring of elementary school's first year the achievement gap in mathematics between whites and blacks widens to -0.635 standard deviations, which is significantly higher than the previous estimates. The score-gap is also widened in reading. However, it differs significantly from the estimate obtained from kindergarten's fall. The relative advantage of Asians lessen, while Hispanics and the other non-white group also show a setback.

3.2 The conditional test score gap

The regression model described by Equation (3.1) does not take into account that other factors also affect the size of the score gap. As family inputs might affect later achievements, we include a set of controls denoted by the vector X in Equation (3.2). This includes: the composite socio-economic status measure (this is time period specific), number of books and it's squares, the mother's age (with dummies for teenager or above 30 years mother), birth weight and family participation in the welfare program *Program for Women, Infants, and Children*. In addition, the student's gender and age are added to the regressions. As we mentioned before, this specification is equivalent to one of the multiple specifications applied by Fryer and Levitt (2004, 2006).

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \text{X}_t' \gamma_t + \varepsilon_i \tag{3.2}$$

The estimation results are included as Table 3.2, where a line separates the racial trait variables from the controls. Implicitly, the higher socio-economic status, age and weight affect positively the expected test scores – all of these are significant both for reading and mathematics in all the three periods. Surprisingly it seems that the effect of family inputs on expected scores diminishes with time. The number of books has a non-linear effect (the squared term is significant). The teenager mother affects negatively the achievements of the children and the mature mother has no significant effect on the scores.

[1]Based on the t-test if the significance level $\alpha = 5\%$ is used.

Table 3.2: Regression results – controls included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	0.116***	0.102***	0.207***	0.003	0.007	0.047*
	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
Black	-0.154***	-0.245***	-0.345***	0.049	-0.033	-0.091***
	(0.024)	(0.026)	(0.027)	(0.025)	(0.026)	(0.027)
Hispanic	-0.222***	-0.188***	-0.272***	-0.094***	-0.139***	-0.201***
	(0.022)	(0.023)	(0.024)	(0.022)	(0.022)	(0.024)
Asian	0.346***	0.191***	0.154***	0.434***	0.268***	0.174***
	(0.036)	(0.034)	(0.032)	(0.044)	(0.038)	(0.033)
Race else	-0.189**	-0.079	-0.171*	-0.170**	-0.146*	-0.112
	(0.067)	(0.069)	(0.079)	(0.060)	(0.063)	(0.083)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.282	0.228	0.237	0.209	0.175	0.171

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.5.

For mathematics, the black and white score gap in kindergarten's fall is narrowed to -0.154 standard deviations when the controls are included. Intriguingly, the score gap is positive when the reading scores are investigated – black students get 0.049 standard deviations more points in reading on average. However, this still implies that the achievement of black and white students with similar family inputs, gender and age is expected to be different. The initial handicap of Hispanic students is also lower, while the advantage of the Asian students is not significantly different when controls are included.

When test scores were remeasured in the spring of kindergarten, the black and white test score gap widened significantly in mathematics even when controls were added. The mathematics scores of black students were 0.245 standard deviations lower on average, while the reading scores of black are 0.007 standard deviations lower on average when controls are included. It is important to emphasize that only the parameter on reading differs significantly from the estimates from kindergarten's fall. As it was observed previously, in the case of the mathematics raw score gap the Hispanic students subdue some of their disadvantage, and the advantage of Asian students became less apparent to the end of kindergarten.

Now let us focus on the test results from the first year of elementary school's spring; the gap between white and black students increased considerably in one year. If we compare two students, one of them is black, while the other one is white and they have the same controlled properties, it is expected that the mathematics score of the black student would be 0.345 standard deviations lower. This means that the average black student did significantly worse than a year before, because the confidence intervals around the point estimates of the black dummy do not overlap. The negative parameter of being black on reading is significantly different from zero – the initial kindergarten advantage of black students in reading disappeared. Black students are the weakest in mathematics when children with the same controlled family inputs are compared. However, they are not significantly weaker than Hispanic students. These findings imply that the cause of the difference might be something else and additional controls should be introduced.

3.3 Test score gap with school fixed effects

A straightforward augmentation of the model is the inclusion of school fixed effects, as ECLS-K samples the students from approximately 800 kindergartens and elementary schools. It is likely that the quality of schools affects the achievements of the students on the tests. Based on the descriptive statistics, the average black student attends an elementary school where 55% of students is black, while the average white student attends an elementary schools where 61% is white. Based on this it can be presumed that the quality of schooling received by the black and white students is intrinsically different.

$$\begin{aligned} \text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \dots \\ + \text{X}_i'\gamma + \text{School}_i'\Theta + \varepsilon_i \end{aligned} \tag{3.3}$$

The inclusion of school fixed effects might capture this difference in schooling quality. The model defined by Equation (3.3) includes school fixed effects – the number of school specific dummies varies as the children move from kindergarten to elementary school. This allows us to take out the average school effect from the achievement of students – essentially we can compare students as they were from the same school or kindergarten.

The estimation results are included in Table 3.3. Interestingly, the size of the gap in different periods regarding mathematics becomes smaller when the school fixed effects are included, but the parameters of being black are not significantly different from the ones that are obtained from the estimation of Equation (3.2). If one would compare a black and white student in the fall of kindergarten from the same school, while these students have the same family inputs that are controlled for, it would be expected that the black student has a 0.160 standard deviations lower mathematics score. This expected achievement gap would be 0.277 standard deviations in the spring of the first elementary school year. This implies that the widening of the achievement gap regarding mathematics starts even when students attend the same school – and the gap nearly doubles in two years. When one turns to the reading scores, it is found that the parameter on being black is negative, but importantly it is not significant. This means that black and white students who are coming from the same school and have the same controlled family inputs are expected to perform similarly in reading in all observed time periods.

Table 3.3: Regression results – controls with school fixed effects

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.160***	-0.240***	-0.277***	-0.050	-0.081*	-0.064
	(0.031)	(0.034)	(0.034)	(0.033)	(0.034)	(0.033)
Hispanic	-0.224***	-0.178***	-0.179***	-0.153***	-0.135***	-0.129***
	(0.026)	(0.028)	(0.030)	(0.026)	(0.027)	(0.030)
Asian	0.267***	0.191***	0.126**	0.366***	0.299***	0.204***
	(0.044)	(0.044)	(0.041)	(0.051)	(0.046)	(0.040)
Race else	-0.124	-0.060	-0.088	-0.062	-0.079	-0.060
	(0.076)	(0.081)	(0.089)	(0.068)	(0.069)	(0.094)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.378	0.360	0.367	0.338	0.330	0.359

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables and school fixed effects are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.6.

A few additional inferences can be drawn from the racial parameters in Table 3.3. Black students lose ground in mathematics on average when we compare their results to the achievements of the Hispanics and Asians from the same school with the same controlled attributes. The initial gap between black and Asian students when they enter kindergarten is 0.347 standard deviations. This is widened to 0.403 standard deviations when the first year of elementary school ends. When one looks at the results of Hispanic children, it comes that they do worse by

0.064 standard deviations when they enter kindergarten. This is turned around when they are in the spring of school's first year – black students have 0.098 standard deviations lower scores than their Hispanic peers.

The results in reading skills show a brighter picture: Asian students perform better than blacks in reading by 0.380 standard deviations when they enter kindergarten. This gap is only 0.268 standard deviations in the spring of elementary school's first year. When the average gap in reading scores between black and Hispanic students is construed, the initial drawback of Hispanics is 0.103 standard deviations. In the next two years this difference is moderated to a 0.065 standard deviation fallback of Hispanic students. In a sum, Hispanic students gain relative to black students while Asian students only gain in mathematics.

The fact that the parameter of the socio-economic status measure became lower when school fixed effects are included implies that students with lower socio-economic status attend lower quality schools on average. The parameters of teenager mother, lack of food security, birth weight and age are not changed significantly by the introduction of the school fixed effects. Importantly, the results in mathematics still do not imply unambiguous segregation or discrimination. Other uncontrolled family inputs, additional school inputs, class effects and unobserved peer effects might cause the difference in achievements (Kertesi and Kezdi, 2011).

3.4 The score gap with school inputs

An alternative augmentation of the model presented in Section 3.2 is the inclusion of school-specific regressors listed in Section 2.2. Adding these explanatory variables to the regression equation results in a model defined by Equation 3.4, which can reveal the aspects of educational facilities contributing to the test score gap.

$$\begin{aligned}\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Asian}_i + \beta_3 \cdot \text{Hispanic}_i + \beta_4 \cdot \text{Race Else}_i + \dots \\ + \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \varepsilon_i \end{aligned} \tag{3.4}$$

Comparing the regression results presented in table 3.4 to the estimates of the model in Section 3.2 we find that controlling for school-specific regressors does not alter the score gap estimates for black children significantly.

Table 3.4: Regression results – controls and school inputs included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	-0.161	0.523*	0.187	-0.544**	0.105	0.161
	(0.215)	(0.224)	(0.227)	(0.179)	(0.234)	(0.257)
Black	-0.152***	-0.229***	-0.347***	-0.034	-0.067*	-0.109***
	(0.029)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)
Hispanic	-0.209***	-0.159***	-0.224***	-0.120***	-0.124***	-0.152***
	(0.025)	(0.026)	(0.027)	(0.025)	(0.025)	(0.027)
Asian	0.305***	0.189***	0.161***	0.366***	0.268***	0.179***
	(0.038)	(0.037)	(0.034)	(0.046)	(0.041)	(0.034)
Race else	-0.177*	-0.055	-0.118	-0.178**	-0.111	-0.063
	(0.069)	(0.070)	(0.080)	(0.061)	(0.064)	(0.083)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School inputs	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.291	0.235	0.211	0.216	0.180	0.178

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables and school inputs are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.7.

A black student having the same family inputs, studying in the same environment as a white student is still expected to score 0.152 standard deviations lower in mathematics in the fall of kindergarten. There is neither in the spring of the kindergarten, nor one year later a significant difference in the widening of the achievement gap. As of the spring of the first year in elementary school, the score gap increases to 0.347 standard deviations. The results imply that regardless the same educational environment, the achievement gap between black and white students in mathematics increases by a factor of two in about two years.

The reading scores are concerning as well. Even though the sign of the score gap is negative in all periods, the estimate on the score gap only becomes significant in the first year of elementary school. That is, selecting a black and a white child with the same family inputs, studying in the same educational environment, it is expected that a significant gap (0.109 standard deviations) in reading skills develops throughout the first year of elementary school.

Regarding the score gap among different minority groups, the results are similar to the regression with school fixed effects. We find that the mathematics score gap between black and Asian students widens from 0.457 standard deviations to 0.508 standard deviations in approximately two years. The reading scores show a significant convergence. The initial 0.4 standard deviations gap narrows to 0.288 standard deviations in the same time period.

The change in the relative performance of black and Hispanic students is subsistent as well. The initial advantage of black children in mathematics in the fall of the kindergarten diminishes, and Hispanic children outperform them by 0.123 standard deviations in the spring of the first year of elementary school.

Regarding the controls, we find that only few of them have a significant and meaningful effect on the performance of the children. One of these significant variables is free lunch, which measures the share of children receiving free lunch in a certain institution, as a general poverty measure of the student body. We find that it may decrease the mathematics score by up to 0.104 standard deviations in the fall of the kindergarten. It has a significantly lower effect (0.063 standard deviations) on the reading scores in the same period. The effect decreases through time, but remains significant in case of mathematics. It is interesting to find that in the fall of the kindergarten the ratio of white students also has a significant negative effect both on mathematics and on reading scores, but this effect is not persistent.

We also find that the per capita number of computers in the educational institution increases the test scores significantly in all periods. If two students of the same race with the same family inputs attend two, otherwise identical facilities, a student in a school having one computer more per capita is expected to score 0.022 standard deviations higher in mathematics, and 0.032 standard deviations higher in reading in the fall of the kindergarten. This effect is the strongest in the spring of the kindergarten with 0.03 and 0.049 standard deviations in mathematics and reading respectively. Even though its impact is decreasing to 0.024 and 0.036 standard deviations by the spring of the first year of the elementary school, it remains significant.

3.5 The score gap with school and family inputs

As noted in 2.2, additional family inputs on the environment of the children are likely to influence their performance, therefore it is worth widening the set of explanatory variables by these variables including book reading, reading time, story telling and whether the family has a computer at home. The exact definition of the variables included is to be found in Appendix B as Table B.1.

Table 3.5: Regression results – controls, school-family inputs included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	-0.395	0.230	-0.155	-0.673***	-0.164	-0.266
	(0.230)	(0.241)	(0.245)	(0.197)	(0.247)	(0.275)
Black	-0.152***	-0.227***	-0.343***	-0.036	-0.066*	-0.107***
	(0.029)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)
Hispanic	-0.194***	-0.143***	-0.204***	-0.103***	-0.101***	-0.128***
	(0.025)	(0.026)	(0.027)	(0.025)	(0.025)	(0.027)
Asian	0.313***	0.198***	0.173***	0.382***	0.286***	0.194***
	(0.038)	(0.037)	(0.034)	(0.046)	(0.040)	(0.035)
Race else	-0.182**	-0.060	-0.121	-0.183**	-0.121	-0.069
	(0.070)	(0.071)	(0.080)	(0.061)	(0.064)	(0.082)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School inputs	Yes	Yes	Yes	Yes	Yes	Yes
Family inputs	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.295	0.239	0.217	0.220	0.188	0.186

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables, school and family inputs are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.8.

The augmentation of the model yields the estimates in Table 3.5. We find that the estimates on racial variables do not change significantly compared to the model without the additional family inputs. As far as the mathematics test scores are concerned, the estimates on being black remain statistically identical. A black student is expected to score 0.152 standard deviations lower in the fall of kindergarten, 0.227 standard deviations lower in the spring of kindergarten, and 0.343 standard deviations lower in the spring of the first year of the elementary school than a white student studying in the same environment with the same family inputs.

As the estimates on other racial variables do not change significantly either, we can still conclude that the racial gap is significant among white and Asian and white and Hispanic children as well. We can observe the decreasing, significant advantage of Asian children in mathematics, the expanding score gap between Asian and black students, and the vanishing initial advantage of black children compared to Hispanic children.

The reading scores, likewise, show a pattern similar to the model without the additional family inputs. In the kindergarten there is no highly significant difference between black and white children, but by the spring of the first year of elementary school a 0.107 standard deviations gap develops. We can still observe the decreasing, significant advantage of Asian children, and the persistent disadvantage of Hispanic children. It still holds, that the difference between black and Hispanic children shows a different pattern in terms of mathematics and reading. In mathematics black children have an initial advantage to Hispanic children which turns around in the spring of the kindergarten, while in reading black children persistently outperform Hispanic children, even though their advantage decreases 0.67 standard deviations to 0.21 standard deviations during the period of our analysis.

The most important school inputs remain the same. We find that the variable Free Lunch still has a significant negative effect on the performance of children, that is, it may decrease the mathematics score by up to 0.099 standard deviations in the fall of the kindergarten. In the same period, it has an up to 0.058 standard deviations negative effect on the reading scores as well. The significance of the variable vanishes in terms of reading by the spring of the kindergarten, but it remains significant in mathematics, even though its magnitude decreases over time. In contrary to the estimates without the additional family inputs, the Public school indicator variable has a significant, 0.078 standard deviations negative effect on the mathematics test scores in the spring of the kindergarten, but it is not significant in any other periods or in terms of reading.

We still find that computers at school are important factors in the performance of the children. Not only the per capita computers at school, but also the presence of a computer at a child's home has a significant effect on test scores. The effect of per capita computers at school remains the same, that is, a one computer increase in the per capita stock of computers at a school is expected to increase the reading scores by 0.031-0.048 standard deviations, and it is expected to increase the mathematics test scores by 0.02-0.028 standard deviations, however, the estimates regarding the mathematics scores are only significant at a 5% level. The computer home variable describing whether a child has a computer at home turns out to be a major factor in the test scores. It has a significant positive effect on both the mathematics and the reading scores in all periods, 0.124-0.138 standard deviations and 0.084-0.156 standard deviations.

All other additional family inputs are insignificant, except for book reading. Book reading is a categorical variable, describing the frequency of parents' reading time. While it has no significant effect on the test scores if the frequency of reading is low, a medium frequency reading is expected to increase the mathematics test scores in the spring of the first year of elementary school by 0.265 standard deviations, and the reading scores in the same period by 0.272 standard deviations. High frequency reading results in a larger positive effect, starting earlier. We find that those children whose parents read frequently are expected to score 0.308 standard deviations higher in mathematics in the spring of the first year of elementary school, and they are also expected to have a 0.265 standard deviations better score in reading in the spring of the kindergarten, which increases to 0.334 standard deviations by the spring of the first year of elementary school.

3.6 The score gap with school and family inputs with fixed effects

The full specification of the model includes the socio-economic status of the children, their school inputs, additional family inputs and school fixed effects. The estimates on the racial regression parameters are presented in Table 3.6, while detailed estimates on all control variables are attached in Appendix B as Table B.9.

Controlling for a wide set of environmental inputs, we find that if two children have the exact same general background, they attend the same school and receive the same school and family inputs, but they belong to different racial groups, the expected achievement gap between them is still significant. The score gap between a black and a white child in terms of mathematics is 0.158 standard deviations in the fall of the kindergarten, which widens to 0.235 standard deviations by the spring of the kindergarten and to 0.308 standard deviations by the spring of the first year of elementary school. It is interesting to find that the reading score gap is only significant starting at the spring of the kindergarten, and only at a 5% level. In the spring of the kindergarten a black child with the same inputs attending the same school is expected to score 0.078 standard deviations lower than a white child in the same kindergarten group, and this score gap remains approximately the same (0.079 standard deviations) one year later as well.

Table 3.6: Regression results – controls, additional inputs with school fixed effects

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.158***	-0.235***	-0.308***	-0.048	-0.078*	-0.079*
	(0.031)	(0.034)	(0.036)	(0.033)	(0.033)	(0.035)
Hispanic	-0.213***	-0.166***	-0.176***	-0.139***	-0.118***	-0.116***
	(0.026)	(0.028)	(0.030)	(0.026)	(0.027)	(0.030)
Asian	0.269***	0.194***	0.128**	0.373***	0.307***	0.211***
	(0.044)	(0.044)	(0.041)	(0.050)	(0.046)	(0.040)
Race else	-0.131	-0.061	-0.088	-0.069	-0.086	-0.062
	(0.078)	(0.081)	(0.089)	(0.070)	(0.069)	(0.092)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School inputs	Yes	Yes	Yes	Yes	Yes	Yes
Family inputs	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.383	0.365	0.373	0.344	0.337	0.366

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables, school and family inputs, as well as school fixed effects are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.9.

In terms of mathematics, the same pattern is present in the case of Asian children as well. The initial better performance of Asian children relative to white children in terms of mathematics narrows during the periods of our analysis from 0.269 to 0.128 standard deviations. The mathematics score gap between black and Asian children is stable, approximately 0.43 standard deviations in all periods. While the mathematics test scores present a similar pattern in case of black and Asian children relative to their white classmates, the pattern in the reading scores is different. In case of the Asian children, the score gap presenting their better performance in reading narrows continuously. Their initial 0.373 standard deviations advantage in reading decreases to 0.307 standard deviations by the spring of the kindergarten, which further decreases to 0.211 standard deviations by the spring of the first year of elementary school.

We can observe a different pattern in terms of mathematics and reading between black and Hispanic children, just like in the models without school fixed effects. The initial advantage of black children in mathematics vanishes by the spring of the kindergarten, and by the spring of the elementary school Hispanic children perform 0.132 standard deviations better. In reading, black children persistently outperform Hispanic children. The score gap between white and Hispanic children in reading is significant even in the fall of the kindergarten, in contrary to the score gap between black and white children. As this score gap narrows from 0.139 standard deviations to 0.116 standard deviations by the spring of the first year of elementary school while the score gap between black and white children widens, we can observe that the score gap of black and Hispanic children gets closer and closer with time.

Regarding family inputs, the most important factor is the socio-economic status of the family, controlled by the SES composite measure. We find that a unit increase in this measure results in a 0.204-0.245 standard deviations better score in mathematics and a 0.217-0.256 standard deviations higher score in reading depending on the period. The number of children's books at home have a significant effect on both mathematics and reading scores. As the square of the number of books is also included in the regression model, we revealed a non-linear relationship between the test scores and the number of books, that is, the positive effect of one additional book to the performance is decreasing in the number of books.

An important result is the estimate on the indicator variable describing the child's gender. In terms of the mathematics scores, we only find a significant difference in the spring of the first year of the elementary school. On a 5% level, male students are expected to have a 0.045 standard deviations lower score in mathematics in the spring of the first year of the elementary school. The difference in terms of reading is, in contrary, significant in all periods. The estimates suggest that male students are expected to score 0.119-0.186 standard deviations higher in reading, with the score gap widening in time.

While the age of the mother lacks significance (only the indicator variable of a young mother is significant on a 5% level), the age of the child when entering school in months is a relevant factor. Even though a child one month older has a 0.226 and 0.166 standard deviations advantage in mathematics and reading respectively in the fall of the kindergarten, this difference narrows to 0.095 and 0.045 standard deviations by the spring of the first year of the elementary school.

The child's weight at birth and the family's participation in the WIC program, which are important possible poverty indicators, also have a significant effect. A child one ounce heavier performs better in mathematics by 0.063-0.072 standard deviations and in reading by 0.04-0.056

standard deviations. Participation in the WIC program results in a 0.117-0 131 and 0.099-0.125 standard deviations lower score in mathematics and reading respectively.

Most effects of the learning environment are represented in the school fixed effects, therefore only a few variables describing the educational facility are significant. The size of class remained significant, and has a positive effect on both mathematics and reading scores, especially in the spring term of the kindergarten. An additional peer in the kindergarten group increases the mathematics and reading scores of the children by 0.068 and 0.082 standard deviations respectively. It is interesting to find that teacher turnover has a positive effect on mathematics performance in the kindergarten. Depending on how severe the turnover is, it increases the test scores by 0.809-0.931 standard deviations in the fall of the kindergarten. Regarding additional family inputs, the computer home variable is the only one that is highly significant. It has a positive effect on both the mathematics and the reading scores in all periods, 0.121-0.14 standard deviations and 0.094-0.152 standard deviations respectively.

3.7 Robustness check

The quantity of data available allows us to analyze whether our estimates are robust to separation of the sample by major factors in the test scores. We built regression models based separation by three major factors, namely the socio-economic status, the ratio of black students in the school and the weight of the child at birth. Regression results are presented in Table 3.7.

Models (1) and (2) present regression estimates of the standardized mathematics test scores in the spring of 2011 on racial and family input variables, separated at the median socio-economic status. We find that the major determinants of test scores in this period are not robust to socio-economic status, that is, different factors influence the performance of children coming from families with above median and below median socio-economic status.

First of all, estimates on racial attributes present a different pattern in the two subsamples, however, the estimates are not different on a 5% level. Therefore we find that belonging to a certain racial group has the same effect on the performance of children statistically, regardless their socio-economic status.

A major finding is that the number of books a child has has a larger influence in case of children with worse socio-economic status. In their case, the estimate on the Books variable is as high as 0.197 standard deviations, while in case of children coming from better environment it is only 0.079 standard deviations. However, the marginal effect of an additional book decreases faster in case of these children, as the estimates on the square of the Books variable are -0.0191 and -0.0035 in case of below median and above median children respectively. An important finding is that the child's weight at birth, as an indicator of general poverty, is more influential in case of below median children as well. It has a 0.07 standard deviations effect compared to the 0.062 standard deviations estimate in case of above median children.

Table 3.7: Regression results – additional robustness checks

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Constant	0.119***	0.282***	0.218***	0.193***	0.176***	0.245***
	(0.032)	(0.026)	(0.027)	(0.029)	(0.031)	(0.031)
Black	-0.319***	-0.359***	-0.279***	-0.374***	-0.337***	-0.361***
	(0.036)	(0.040)	(0.079)	(0.032)	(0.036)	(0.040)
Hispanic	-0.269***	-0.225***	-0.271***	-0.270***	-0.282***	-0.258***
	(0.032)	(0.038)	(0.033)	(0.034)	(0.035)	(0.033)
Asian	0.187***	0.139***	0.158***	0.151***	0.176***	0.123*
	(0.055)	(0.040)	(0.046)	(0.045)	(0.042)	(0.049)
Race else	-0.230*	-0.073	-0.088	-0.399**	-0.116	-0.230
	(0.092)	(0.136)	(0.095)	(0.121)	(0.093)	(0.128)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	6843	6745	7113	6475	6806	6782
R^2	0.110	0.146	0.200	0.196	0.178	0.221

Notes: The columns report weighted OLS regression results. The standardized mathematics test score in the spring of 2011 is the dependent variable. The racial attribute dummies and standardized baseline control variables are used as regressors. Robustness checks (1) and (2) are regarding children who have below and above the median socio-economic status values. Specifications (3) and (4) are separated based on the below and above the median ratio of black students in the school. Models (5) and (6) are estimated on the subsample of below and above median weight children. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The detailed version is enclosed in Appendix B as Table B.10.

Models (3) and (4) are estimates of the standardized mathematics test scores in the spring of 2011 on racial and family input variables, separated at the median ratio of black students in the school. In this case, we encounter robust estimates, that is, the ratio of black children in a school does not make a statistically significant difference between the major factors in children's test scores.

The same holds for the weight of the child at birth, with two exceptions. Models (5) and (6) suggest that the weight at birth is indeed, a likely indicator of poverty. We find that children with above median weight at birth in general have an advantage, as the estimate on the constant in their case (0.245 standard deviations) is significantly higher than in case of below median children (0.176 standard deviations). All other estimates are statistically identical in the two subsamples, except for participation in the WIC program, which is an other indicator of poverty in the sample. We find that those who have above median birth weight, that is, those who are likely to come from a family with better financial status, are more sensitive to the participation in the WIC program. In case of children above median, participation in the WIC program has a -0.2 standard deviations effect, while in case of below median children it only has a -0.128 standard deviations effect. That is, we find an advantage of 0.069 standard deviations of children above median birth weight in general, which is corrected by a -0.072 standard deviations effect from the difference on the estimates on the WIC program, equating each other statistically. In general, we find that the child's socio-economic status makes a difference in the major factors determining the test score, but ratio of black children and birth weight do not.

3.8 Comparison with an earlier cohort

The estimates in Tables 3.1, 3.2 and 3.3 are based on the same model specifications as estimates of Fryer and Levitt (2004). Their research is also based on ECLS-K, but it is about the cohort of students who entered kindergarten in 1998. About the initial raw score gap between black and white students: regarding mathematics there was a -0.638 standard deviations gap in the fall of 1998, when students entered kindergarten. The unconditional gap was -0.401 standard deviations in reading. These results would imply that the 2010 cohort of black students gained both in mathematics and reading relative to those who entered kindergarten in 1998.

When the controlled results of Table 3.3 are investigated, the coefficients obtained by Fryer and Levitt (2004) are lower in absolute terms. In the fall of kindergarten the gap regarding mathematics was only -0.094 standard deviations. Later in the spring of kindergarten it increased to -0.201 standard deviations, in the spring of the first school year it was -0.250 standard deviations. This also implies that the widening time pattern of the mathematics score gap is unchanged after 12 years.

Importantly, the estimated coefficients of being black regarding the fixed effects models are not significantly different than the parameters describing the 1998/1999 relationship. The coefficients of reading became smaller in absolute terms – which might imply that the fallback of blacks in this regard became smaller if students from the same schools are investigated. In mathematics, the initial handicap of black students was smaller initially in 2012, but to the end of elementary school's first year the setback of black students was larger according to our estimates than it was in 1998.

Chapter 4: Counterfactual approaches

This chapter investigates the black and white score gap with advanced methodologies. In Section 4.1 we give a general overview about the achievement of black and white students. We take a close look at those environmental factors that are associated with lower and higher achievement. The score gap that emerges among black and white students is broken up into different effects with Oaxaca-Blinder decomposition in Section 4.2. We do a similar counterfactual investigation in Section 4.3, where we applied racially matched linear regressions. The findings of the counterfactual methods are consistent, both of the models imply that the initial black and white score gap is present among students because of environmental factors. At the same time, we also found supporting evidence that the widening of the gap is due to heterogeneous returns to family and school inputs among blacks and whites.

4.1 Black and white specific models

For the purpose of Blinder-Oaxaca decomposition we estimate equation 4.1 separately for black and white students, with the previously introduced notation holding. We must note, that in order to be able to perform the decomposition, we estimate the regressions on level in this case.

$$\text{Test Score}_i = \beta_0 + \mathrm{X}_i'\gamma + \mathrm{Y}_i'\Upsilon + \mathrm{Z}_i'\Gamma + \varepsilon_i \tag{4.1}$$

The estimation results are in Appendix B as Tables B.11 and B.12. Estimating effects on level instead of estimating them on standardized test scores yields a different interpretation of the estimates. In this case, the coefficient of a certain variable means that holding all other controlled variables constant, a unit increase in the variable in question results in a change in the dependent variable given by the estimated coefficient.

We ran two separate regressions on Black and White children, containing the same explanatory variables. The Black race specific model's estimates are attached as Table B.11 in Appendix B. We find that in case of Black children, the socio-economic status of the family has a major effect on the test scores, both in mathematics and in reading, in all periods. We find that a unit increase in the composite socio-economic status measure increases the conditional expectation on the mathematics test score of the child by 3.946-4.443 points, depending on the period. In case of reading, this effect is 3.592-4.901 points. It can be observed, that the magnitude of this effect increases by time in both cases, that is, the socio-economic status has an increasing importance as the child continues their studies.

We find that the number of children's books is important in the early performance of the children. In the fall of the kindergarten, it has a significant positive, but decreasing marginal effect on the mathematics score. In terms of reading this effect is significant in the spring of the kindergarten as well.

As far as gender is concerned, being male has a significant positive effect on the reading scores, but not on the performance in mathematics. A Black boy is expected to have a 0.983 points higher reading score than a Black girl in the fall of the kindergarten, if they are identical in the other controlled inputs. This score gap increases to 2.176 and 2.962 points by the spring of the kindergarten and the spring of the first year of elementary school respectively.

The child's age also matters. Starting the school one month older results in a 0.542 points increase in the mathematics score, and a 0.393 points increase in the reading score. Even though the magnitude of this effect decreases with time, it remains significant with 0.291 and 0.188 points in mathematics and reading respectively in the spring term of the first year of the elementary school.

We find that weight at birth and the age of the mother is not significant, but participation in the WIC program has a negative effect on both the mathematics and the reading scores. Considering two Black kids with the exact same inputs except for the participation in the program, the one participating is expected to score 1.657-2.001 points lower in mathematics and 1.689-2.456 points lower in reading.

Regarding the white race specific regression model, the socio-economic status has an important role as well. Holding other factors constant, a unit increase in the composite SES measure results in a 4.449-4.977 points expected increase in the mathematics score and a 3.601-5.021 points expected increase in the reading score. The number of children's books have a significant positive, but diminishing marginal effect as well in all periods, both in mathematics and in reading. Just like in the black race specific regression model, gender has a significant effect on the reading scores. A male student is expected to have a 1.255 points higher reading score than a female student in the fall of the kindergarten, if they are identical in the other controlled inputs. The score gap is even higher in the spring of the kindergarten and the spring of the first year of elementary school, 1.878 and 2.543 points respectively.

We find that the age of the mother does not have a significant effect, but the age of the child has. Starting the school one month older results in a 0.65 points increase in the mathematics score, and a 0.434 points increase in the reading score in the fall of the kindergarten. The magnitude of this effect decreases with time just like in the black specific regression. In the spring term of the first year of the elementary school it is 0.291 and 0.188 points in mathematics and reading respectively. The child's weight at birth and the family's participation in the WIC program also have a significant effect. A child one ounce heavier performs better in mathematics by 0.047-0.054 points and in reading by 0.025-0.049 points. Participation in the WIC program results in a 1.771-2.499 and 1.378-2.416 points lower score in mathematics and reading respectively.

4.2 Blinder-Oaxaca decomposition

The conditional gap in the achievements can be decomposed into the three effects that were described in Section of Chapter 2. These effects namely are the following: difference in endowments, differences in sensitivity and the difference caused by the interaction of the previous two.

The black and white race specific regressions' results that were discussed in Section 4.1 and enclosed in Appendix B as Tables B.11 and B.12 can be used for implementing an Oaxaca Blinder decomposition. The decomposition can be done both for mathematics and reading in three distinct time periods. Students of other races and students with mixed racial backgrounds were excluded from the estimation. The period and subject specific results, the controlled difference estimates and decomposition results are enclosed in Table 4.1.

Let us first discuss the difference in mathematics, the results presented in Table 4.1 are on the levels as we had previously emphasized this. In the fall of 2010 the difference in mathematics was -2.055 points. This difference among black and white students is solely attributable to the difference in endowments (socio-economic status, lower quality schools, lack of racial mixing in schools). The other parts of the counterfactual decomposition are insignificant – this implies that black and white students who have the exact same attributes are doing similarly in terms of mathematics in the beginning. The school clustered standard errors are part of a very stringent research set up – these results show that we can state with an extremely high confidence that the initial existence of the black and white students is due to environmental factors. Not a whole academic year passes by until the next measurement period – interestingly the difference became wider. This widening is also significant in statistical terms – at $\alpha = 5\%$ the differences are different from each other, black students do even worse than they did previously. Approximately the increase in fallback is roughly a whole test score point. Interestingly, the decomposition results are more informative than they were previously. The coefficient effect (the sensitivity effect) became significant, and negative. Black students do not receive the same returns to family and school inputs that white students receive. Importantly the size of endowments effect is unchanged – the fallback is only due to the changing sign of the coefficients effect. The results of the second year show a similar picture – the difference increases roughly by whole point, and the difference is solely attributable to the change coefficients.

Table 4.1: Blinder-Oaxaca decomposition results

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
White	30.876***	43.989***	63.87***	37.49***	49.84***	69.90***
	(0.119)	(0.124)	(0.137)	(0.103)	(0.124)	(0.135)
Black	28.811***	40.982***	59.88***	37.11***	48.81***	68.30***
	(0.200)	(0.221)	(0.237)	(0.183)	(0.227)	(0.241)
Difference	-2.055***	-3.007***	-3.984***	-0.381	-1.023***	-1.602***
	(0.233)	(0.253)	(0.274)	(0.210)	(0.258)	(0.277)
Endowments	-2.307***	-2.201***	-2.285***	-1.674***	-1.842***	-1.991***
	(0.134)	(0.129)	(0.131)	(0.106)	(0.120)	(0.128)
Coefficients	0.374	-0.762**	-1.952***	1.458***	1.065***	0.164
	(0.248)	(0.269)	(0.289)	(0.242)	(0.304)	(0.290)
Interaction	-0.122	-0.044	0.253	-0.165	-0.247	0.225
	(0.155)	(0.177)	(0.185)	(0.146)	(0.197)	(0.189)
N	12222	12222	12222	12222	12222	12222

Notes: The columns report the Blinder-Oaxaca decomposition results. Mathematics and reading unstandardized test scores of students in different time periods are the dependent variables. The regression results enclosed in Appendix B as Tables B.11 and B.12 were used for the calculations. The difference is decomposed into endowments, coefficients and interaction effects in the rows. School clustered robust standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In case of writing the decomposition of reading results shows that the initial difference among black and white students in the fall of 2010 is negative and insignificant. Parts of the three decomposition elements are significant: the endowments effect is negative and significant, while the coefficients effect is positive. This essentially demonstrates that black students who have similar background as whites are expected to outperform their white peers in reading. This is turned around within one year, the disadvantage of black students became significant to the end of the first year. On level the controlled reading score gap was one point among black and white students in the spring of 2011. The trajectory was unchanged, the handicap increased by a half score to the end of the first school year. As before the increased fallback was due to the coefficients effect. However, this change was not significant in a year-to-year comparison. The black and white reading score gap was not that severe historically.

The coefficients (sensitivity) part of the Oaxaca-Blinder decomposition gives a higher estimate of possible differences that are due to discrimination. However, we would caution against considering only this interpretation of our findings, because the existence of significant coefficient effects might flag multiple important phenomena separately or a mixture of them. The main possible effects behind the observed divergence of coefficients effect can be summarized as:

1. The tests in later periods have increasingly parts that would need systematic studying skills. If black students are in lack of such skills their results would definitely fall behind white students. *It should be emphasized, that under this scenario an output based incentive program targeting black students is ineffective.*

2. The absolute level of dispersion in scores increases, previously insignificant differences among black and white students might become significant. This would imply a similar phenomenon to the one that Brainerd (2000) describes about the evolution of the gender wage gap in post-socialist countries.

3. The black students are victims of premarket discrimination, they attend worth schools where the teachers are also lower quality. Initially they are just behind because of their disadvantaged family background, but later on they fall back more because of lower quality schools. Another form of premarket discrimination would be if teachers would increasingly and explicitly discriminate black students and downgrade their efforts.

4. Unmeasured factors are affecting the achievement of black students, which drive the emergence of the significant coefficients effect in case of mathematics. Also this factor

might be latently behind the disappearance of the positive coefficients effect in case of reading.

4.3 Propensity score matching

As we previously pointed out in an ideal experimental research setting the black and white race of students would be assigned randomly to students. However, this ideal setting will never happen, but we can imitate such research settings with propensity score matching. We used nearest neighbor propensity score matching to pair up black and white students who are similar in environmental factors such. Variables included in the matching are The complete list of variables is the set of environmental controls, the family and school input variables. With this method we generated an artificial sample of students. In the matching we sampled with replacement, because we allowed for pairing multiple black students to the same white peer. Our unique matched sample is used for the estimation of Equation (4.2) for mathematics, reading in all three periods. A summary table of the results is enclosed in the body of the text as Table 4.2 the main table with detailed results is in Appendix B as Table B.13.

$$\text{Test Score}_i = \beta_0 + \beta_1 \cdot \text{Black}_i + \text{X}_i'\gamma + \text{Y}_i'\Upsilon + \text{Z}_i'\Gamma + \varepsilon_i \tag{4.2}$$

In our matched sample black students do not perform worth significantly than whites in the beginning of kindergarten. They have a slight disadvantage – around 0.032 standard deviations, but the standard errors are nearly double the size of the black coefficient. A number of controls has lost it's significance in the regression including: the WIC dummy, the birth weight and the specific school and family inputs. In the second term of the school year black students fall behind the white ones significantly, their disadvantage is suddenly -0.145 standard deviations. This is quite shocking if one takes into account the fact that because of the matching these black and white students have the similar social background, in certain aspects similar schools and similar quality of family inputs. This finding underpins the findings of Section 4.2, where we found that the divergence starts in the formal education system. This suggests that a policy which deals with shrinking the black and white score gap has to happen as early as kindergarten. To the end of the first school year the mathematics gap nearly doubles, this means that the general trajectory is unchanged – black students lose constantly against white in mathematics even when similar students are compared.

In the matched regression with mathematics score dependent variable the role of socio-economic status is reduced. The children's initial age also loses from it's importance, and also the disadvantage of children who have a young mother decreases with time. The partial effect of books on the children's expected achievement is concave as it was before. It should be also

Table 4.2: Regression results – score gap after propensity score matching

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	0.651	0.696	0.313	-0.789	-0.340	0.281
	(0.737)	(0.490)	(0.522)	(0.419)	(0.425)	(0.518)
Black	-0.032	-0.145*	-0.319***	0.073	-0.031	-0.085
	(0.053)	(0.061)	(0.054)	(0.046)	(0.057)	(0.056)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
School inputs	Yes	Yes	Yes	Yes	Yes	Yes
Family inputs	Yes	Yes	Yes	Yes	Yes	Yes
N	4076	4076	4076	4076	4076	4076
R^2	0.247	0.183	0.114	0.167	0.156	0.120

Notes: Columns report OLS regression results after propensity score matching. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The racial attributes, controls, family and school inputs are used as regressors, while multiracial students are excluded from the sample. Clustered standard errors are included in the parentheses. The regression results are enclosed in Appendix B as Table B.13. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

noted that the fit of the models is becoming weaker – the unobserved heterogeneity in the models increases with time. In other words a part of the score gap is due to unobserved effects, and this part is increasing.

Based on the matched regressions, black students do better than whites in reading by 0.073 standard deviations in the fall of kindergarten, but this advantage is not significant. Also, the fit of the model from the first period is significantly lower than the fit of the mathematics specific model. The school inputs are not significant covariates, the role of the family related environmental factors is similar as it was before. The initial advantage of black students in reading disappears even in the controlled regression, their relative disadvantage is -0.030 standard deviations. This means that within one years, they lose ground considerably – their loss is roughly 0.103 standard deviations. The fit of the reading specific model is decreased (more unobserved heterogeneity is present) and the initial advantage of students who entered kindergarten at a higher age starts to lessen. By the end of the first elementary school year the disadvantage of black students is prevalent, they are losing ground in reading strongly – they are behind white students by 0.085 standard deviations in reading. This is not significant, but these results also support the ones tat were obtained by the use of decomposition methods in Section 4.2.

Chapter 5: Conclusion

In this final chapter we summarize our findings. In Section 5.1 we give detailed answers to the research questions posed in Chapter 1. We also support our argument with plots that allow for the cross model inspection of our estimates. Possible extensions of our current research are sketched out in Section 5.2 in detail.

5.1 Main findings

The paper in hand investigated five closely related questions about the black and white score gap among kindergarten and elementary school students. These questions were answered by our empirical investigation with a fair confidence. The backbone of this investigation was a set of linear regressions which we supported by counterfactual approaches. Our proposed answers to the research questions can be summarized as:

1. *How large was the achievement gap among black and white kindergarten and elementary school students who started preschool in 2010/2011?* The uncontrolled mathematics gap is around 0.5 standard deviation in all observed periods. On Figure 5.1 one can see that the controlled mathematics score gap starts at the approximate value of 0.2 standard deviation and this increases to 0.3 standard deviations. The reading scoregap is plotted on Figure 5.2. The unconditional gap among black and white students is approximately 0.3 standard deviation in the fall of kindergarten, this increases to 0.4 standard deviations. Initially the conditional gap with controls is insignificant, later black students lose ground and the gap is around 0.1 standard deviations.

2. *What are the possible causes of the achievement gap?* The results of ordinary least-squares regressions and the findings of counterfactual methods show that the initial achievement gap is mainly rooted in the differences in endowments. These are namely the differences in the children's socio-economic status, school quality, parental inputs and school inputs. The later divergence in the scores is caused by the emergence of heterogeneous re-

turns to endowments among racial groups according to the results of the Oaxaca-Blinder decomposition.

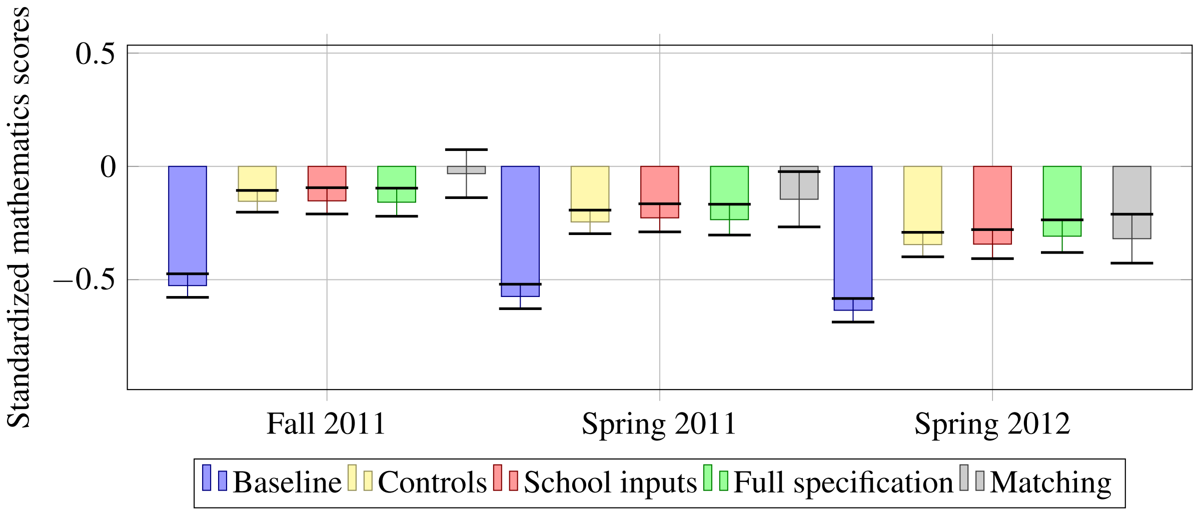

Figure 5.1: Estimates of the mathematics score gap

3. *Under which conditions does the gap disappear?* The initial gap disappears if black and white students with the same socio-economic background are compared to each other – this condition stands if one uses endowment–coefficient effect decomposition or propensity score matching. The mathematics gap remains significant in the second and third periods even in a counterfactual setting – see Figure 5.1.

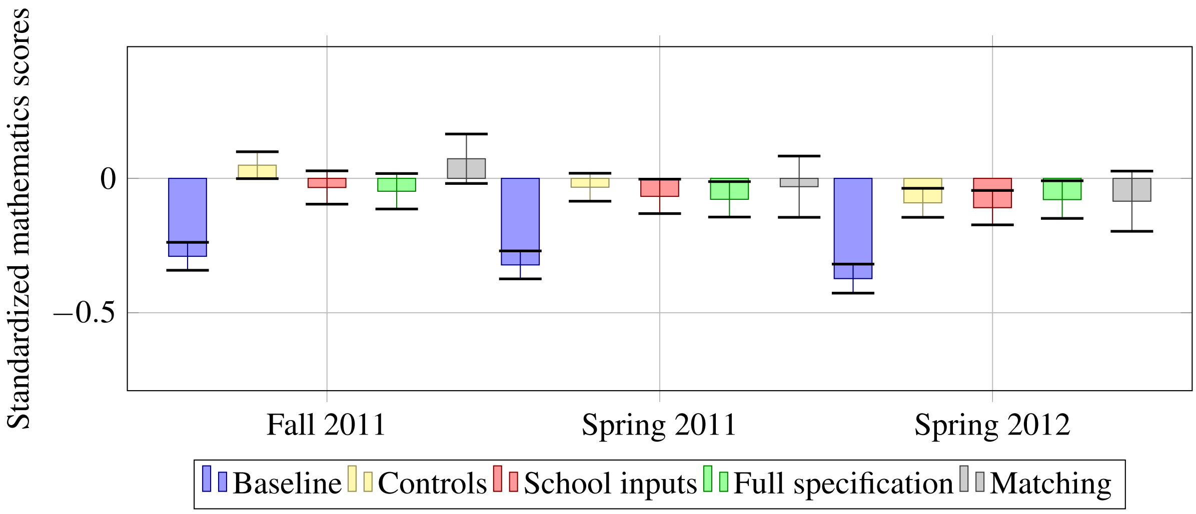

Figure 5.2: Estimates of the reading score gap

4. *Is the black and white test score gap became smaller or greater since 1998?* The sizes of the mathematics and reading gaps are not significantly (in a statistical sense) different

from the ones that Fryer and Levitt (2004) obtained. This is true in all of the specifications that we apply: the baseline one, the socio-economic status controlled one and for the model with school fixed effects. Intriguingly, this is also true for the models that use additional family and school input controls

5. *Was the No Child Left Behind effective in terms of closing the relative gap?* The results solely show that the achievement gap remained significant among black and white students. In addition, the trajectory of the gap is the same – as terms go by the size of the relative gap increases both in mathematics and verbal skills. Importantly the size of the gap is unchanged, so we conclude that the NCLB was not effective universally in reducing the relative gap. We have to note that this does not mean that black students might have not experienced absolute gains regarding the skills in focus.

5.2 Further extensions

The research that we done can be extended in multiple genuine ways. We consider mainly two possibilities – extending the research by introducing new variables and the methodological improvement of our work. The extension of the research with new variables can be summarized as follows:

1. Our specifications were in line with the ones used by (Fryer and Levitt, 2004) because we made comparisons among cohorts of the ECLS-K. This meant that the set of predictors that we used was constrained to a handful of variables. However, the ELCS-K dataset is rich in possible controls – the parents' background can be extended regarding their education level, drug use and previous employment status. Additional school inputs might be also included in the regressions to control for school quality.

2. The policies that were part of NCLB were not introduced by every school in the United States. The effect of these policies can be investigated if schools were not self-selecting themselves into implementing the policies. The schools are identifiable by unique identification codes in the dataset and matching up with external datasets about partial implementation of the policies is possible because of this. Through this the effect of the different policy elements can be estimated.

3. The school fixed effects can be biased if certain schools have multiple classes in the sample. This omitted variable bias might also effect the size of the being black parameter. If the schools that have multiple classes in the sample segregate based on race the school

fixed effect is possibly useless. However, the large number of class fixed effects would make the computation of estimates extremely demanding.

4. The schools might also experience a strong homophily based peer effect, as Jackson et al. (2016) points out the race based homophily might be correlated with other types of homophilies (e.g. homophily in income, parents' education or gender). This phenomenon might imply that the estimates of the black and white student ratio parameters are biased, due to endogeneous homophily measures. Controls that measure homophily and assortative mixing in other variables might change the sign and magnitude of the above mentioned ratio parameters.

The followings are possible methodological extensions of our work that can serve for the base for future research on the black and white test score gap:

1. There are decomposition methods that can deal with explaining the sources of variance in an outcome variable among groups in time such as the Juhn-Murphy decomposition (Junh et al., 1991). Applying these methods to the time dimension would allow for a better understanding of the trajectory of the black and white score gap.

2. Our approach only applies of one single propensity score matching method. However, the literature has numerous propensity score matching methods that might result in qualitatively different results (Ho et al., 2007). A robustness testing of our results would be also a neat extension of our work. Propensity score matching methods that might be considered can include: multiple neighbors, distance based, semi-supervised or kernel based.

3. The scores from the spring term might be used for panel data econometrics – the equal time difference among the sampling periods allows the use of various panel data methods. Importantly, the existence of school identifiers makes possible the estimation of multi-dimensional panel data models. In such models the additional dimension (above individuals and time) could be the school or class of students.

4. From the first school year (this is not true for kindergarten) in fall periods only a subsample of students is tested, this means that the dataset would become strongly unbalanced. Imputation with chained equations for the fall periods makes the cross-time comparison more detailed – the higher frequency measurement might identify important breaking points in the divergence of scores.

Bibliography

Andrisani, P. and Daymont, T. (1984). Job Preferences, College Major, and the Gender Gap in Earnings. *Journal of Human Resources*, 19:408–428.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Azur, M. J., Stuart, E., Frangakis, C., and Leaf, P. (2011). Multiple Imputation by Chained Equations: What is it and How Does it Work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4):436–455.

Bradley, A. and Fryer, R. (2011). *The Power and Pitfalls of Education Incentives*. Brookings Institution, Hamilton Project.

Brainerd, E. (2000). Women in Transition: Changes in Gender Wage Differentials in Eastern Europe and the Former Soviet Union. *Industrial and Labor Relations Review*, 54(1):138–162.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

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

Darling-Hammond, L. (2007). Race, Inequality and Educational Accountability: the Irony of No Child Left Behind. *Race, Ethnicity, and Education*, 10(3):245–260.

Dee, T. S. and Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management*, 30(3):418–446.

Franklin, W. (1999). *Immigration and Opportunity: Race, Ethnicity, and Employment in the United States*, chapter Ethnic Concentrations and Labor-Market Opportunities, pages 106–140. Russel Sage Foundation.

Fryer, R. (2011). *Handbook of Labor Economics*, volume 4, chapter Racial Inequality in the 21st Century: the Declining Significance of Discrimination, pages 855–971. Elsevier.

Fryer, R. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*, 31(3):373–427.

Fryer, R. and Dobbie, W. (2011). High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3):158–187.

Fryer, R. and Levitt, S. (2004). Understanding the Black and White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, 86(2):447–464.

Fryer, R. and Levitt, S. (2006). The Black-White Test Score Gap Through Third Grade. *American Law and Economics Review*, 8(2):228–249.

Fusarelli, L. D. (2004). The Potential Impact of the No Child Left Behind Act on Equity and Diversity in American Education. *Educational Policy*, 18(1):71–94.

Gaddis, M. and Lauen, D. L. (2014). School Accountability and the Black–White Test Score Gap. *Social Science Research*, 44:15–31.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag.

Ho, D., Imai, K., King, G., and Stuart, E. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 15(3):199–236.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Jackson, M. O., Rogers, B., and Zenou, Y. (2016). The Economic Consequences of Social Network Structure. Available at SSRN: http://ssrn.com/abstract=2467812 or http://dx.doi.org/10.2139/ssrn.2467812.

Junh, C., Murphy, K., and Pierce, B. (1991). *Workers and Their Wages*, chapter Accounting for the Slowdown in Black-White Wage Convergence, pages 107–143. AEI Press.

Kertesi, G. and Kezdi, G. (2011). The Roma/Non-Roma Test Score Gap in Hungary. *American Economic Review*, 101(3):519–525.

Kertesi, G. and Kezdi, G. (2015). On the Test Score Gap Between Roma and Non-Roma Students in Hungary and its Potential Causes. *Economics of Transition*, 23(2):1–28.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Quinn, D. M. (2015a). Black–White Summer Learning Gaps Interpreting the Variability of Estimates Across Representations. *Educational Evaluation and Policy Analysis*, 37(1):50–69.

Quinn, D. M. (2015b). Kindergarten Black–White Test Score Gaps: Re-examining the Roles of Socioeconomic Status and School Quality with New Data. *Sociology of Education*, 88(2):120–139.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Appendix A: Notations

β_0 – Intercept parameter

β_1 – Parameter of being black

β_2 – Parameter of being Asian

β_3 – Parameter of being Hispanic

β_4 – Parameter of being Pacific Asian or Native American

γ – Vector of parameters regarding the control variables

Ω_B – Vector of parameters regarding the black specific regression

Ω_W – Vector of parameters regarding the white specific regression

Θ – Vector of parameters regarding the school fixed effects

Υ – Vector of parameters regarding school inputs

Γ – Vector of parameters regarding family inputs

η – Vector of parameters in the matching probit model

ε – Idiosyncratic error term

Φ – Cumulative distribution function of the standard normal distribution

X – General control variables

Y – School input variables

Z – Family input variables

W – Controls used for matching

Appendix B: Tables

Table B.1: Description of variables used in the analysis – compiled by the authors

Variable	Description
Reading score	Aggregated reading score in the given period
Mathematics score	Aggregated mathematics score in the given period
Asian	Asian dummy: 1 if Asian; 0 otherwise
Black	Black dummy: 1 if black; 0 otherwise
Hispanic	Hispanic dummy: 1 if hispanic; 0 otherwise
White	White dummy: 1 if white; 0 otherwise
Race else	Race else dummy: 1 if Native American or Pacific Asian; 0 otherwise
Books	Number of children's books in home
Child's age	Age of child in month's on entering school
SES	Socio-economic status composite measure
Gender	Gender dummy: 1 if male; 0 otherwise
Weight	Birthweight in ounces
WIC	WIC dummy: 1 if participant of *Program for Women, Infants, and Children*
Young mother	Teenage mother dummy: 1 if mother was under 18 when child was born
Mature mother	Mature mother dummy: 1 if mother was above 30 when child was born
Free lunch	Share of students receiving a free lunch in the school/kindergarten
Public school	Public school dummy: 1 if school is a public school; 0 otherwise
Ratio of whites	Share of white students in the school/kindergarten
Ratio of blacks	Share of black students in the school/kindergarten
Turnover	Categorical variable describing teacher turnover
Security	Security dummy: 1 if students have to pass a security gate; 0 otherwise
Internet access	Internet access dummy: dummy: 1 if school has internet; 0 otherwise
Computers	Per capita computer in school/kindergarten
Drugs	Categorical variable describing drug problems around school
Gang problems	Categorical variable describing gang problems around school
Class size	Number of students enrolled in the class
Book reading	Categorical variable describing the frequency of parents' reading time
Reading time	Variable describing length of parents' reading time
Story telling	Categorical variable describing the frequency of parents' story tellings
Computer home	Owning computer dummy: 1 if student has a computer; 0 otherwise
School FE	Dummies describing the schools

Table B.2: Abbreviations used in the paper

ABBREVIATION	THE ABBREVIATED EXPRESSION
ECLS-K	Early Childhood Longitudinal Program – Kindergarten
NCLB	No Child Left Behind Act
MICE	Multiple Imputation by Chained Equations
PSM	Propensity Score Matching
SES	Socio-Economic Status
WIC	Program for Women, Infants, and Children

Table B.3: Average test results by race

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Asian	35.942	47.909	67.449	42.406	54.591	74.114
	(11.628)	(10.970)	(12.128)	(13.210)	(14.040)	(11.909)
Black	26.916	38.972	56.276	35.89	47.095	66.018
	(9.419)	(10.823)	(12.238)	(8.471)	(11.316)	(13.271)
Hispanic	25.202	39.005	56.222	33.865	46.012	63.934
	(9.766)	(11.340)	(12.630)	(8.055)	(11.573)	(13.916)
White	30.996	44.145	63.889	37.448	49.809	69.849
	(10.939)	(11.395)	(13.265)	(9.341)	(11.491)	(13.269)
Race else	27.798	42.181	60.832	35.048	47.928	67.759
	(10.321)	(11.278)	(12.552)	(9.307)	(11.956)	(13.596)
Mean	30.564	43.568	62.944	37.463	49.714	69.51
	(10.989)	(11.518)	(13.375)	(9.688)	(11.867)	(13.356)

Standard errors in parentheses.

Table B.4: Mean of controls by race in the observed time periods

	Asian	**Black**	**Hispanic**	**White**	**Race else**
SES 2010 fall	0.334	-0.370	-0.524	-0.016	-0.280
	(0.025)	(0.015)	(0.011)	(0.007)	(0.046)
SES 2012 spring	0.266	-0.387	-0.555	-0.007	-0.282
	(0.026)	(0.019)	(0.012)	(0.008)	(0.044)
Weight	111.177	110.316	116.395	117.883	117.209
	(0.710)	(0.524)	(0.409)	(0.216)	(1.626)
Ratio of whites	38.409	25.412	30.064	62.497	44.262
	(0.576)	(0.629)	(0.519)	(0.324)	(1.931)
Ratio of blacks	9.678	55.012	10.684	8.308	6.424
	(0.344)	(0.500)	(0.262)	(0.128)	(0.927)
Size of class	20.940	20.246	20.793	20.008	19.903
	(0.440)	(0.219)	(0.208)	(0.125)	(0.588)
Child's age	66.157	67.406	66.913	67.649	67.528
	(0.130)	(0.100)	(0.075)	(0.044)	(0.313)
Free lunch	34.355	61.162	60.296	39.488	49.002
	(0.768)	(0.662)	(0.489)	(0.288)	(1.928)
Mother's age	36.069	32.406	32.730	34.287	32.429
	(0.177)	(0.179)	(0.128)	(0.068)	(0.517)
Books	62.316	44.865	50.989	96.012	60.231
	(3.23)	(1.365)	(2.275)	(1.497)	(6.978)

Notes: The columns report the race specific means. Multiracial students are excluded from the sample. Standard errors are included in parentheses.

Table B.5: Regression results – controls included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.154***	-0.245***	-0.345***	0.049	-0.033	-0.091***
	(0.024)	(0.026)	(0.027)	(0.025)	(0.026)	(0.027)
Hispanic	-0.222***	-0.188***	-0.272***	-0.094***	-0.139***	-0.201***
	(0.022)	(0.023)	(0.024)	(0.022)	(0.022)	(0.024)
Asian	0.346***	0.191***	0.154***	0.434***	0.268***	0.174***
	(0.036)	(0.034)	(0.032)	(0.044)	(0.038)	(0.033)
Race else	-0.189**	-0.0786	-0.171*	-0.170**	-0.146*	-0.112
	(0.067)	(0.069)	(0.079)	(0.060)	(0.063)	(0.083)
SES	0.326***	0.294***	0.250***	0.319***	0.285***	0.264***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Books	0.110***	0.099***	0.111***	0.106***	0.092***	0.082***
	(0.016)	(0.015)	(0.015)	(0.017)	(0.015)	(0.014)
Books^2	-0.0052***	-0.0046***	-0.0052***	-0.0050***	-0.0044***	-0.0037***
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0009)	(0.0010)
Gender	0.019	0.033	-0.022	0.131***	0.167***	0.212***
	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)
Child's age	0.247***	0.181***	0.114***	0.194***	0.126***	0.067***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Weight	0.076***	0.079***	0.067***	0.042***	0.055***	0.060***
	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Young mother	-0.100*	-0.074	-0.082	-0.080*	-0.030	-0.099
	(0.047)	(0.052)	(0.056)	(0.040)	(0.049)	(0.055)
Mature mother	0.020	0.012	-0.007	0.008	-0.029	-0.020
	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
WIC	-0.168***	-0.136***	-0.164***	-0.150***	-0.114***	-0.175***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Constant	0.116***	0.102***	0.207***	0.003	0.007	0.047*
	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
N	13588	13588	13588	13588	13588	13588
R^2	0.282	0.228	0.237	0.209	0.175	0.171

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.6: Regression results – controls and school fixed effects

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.160***	-0.240***	-0.277***	-0.050	-0.081*	-0.064
	(0.031)	(0.034)	(0.034)	(0.033)	(0.034)	(0.033)
Hispanic	-0.224***	-0.178***	-0.179***	-0.153***	-0.135***	-0.129***
	(0.026)	(0.028)	(0.030)	(0.026)	(0.027)	(0.030)
Asian	0.267***	0.191***	0.126**	0.366***	0.299***	0.204***
	(0.044)	(0.044)	(0.041)	(0.051)	(0.046)	(0.040)
Race else	-0.124	-0.060	-0.088	-0.062	-0.079	-0.060
	(0.076)	(0.081)	(0.089)	(0.068)	(0.069)	(0.094)
SES	0.259***	0.249***	0.217***	0.268***	0.263***	0.233***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Books	0.098***	0.087***	0.095***	0.110***	0.084***	0.060***
	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.014)
Books^2	-0.0047***	-0.0041***	-0.0045***	-0.0051***	-0.0038***	-0.0026**
	(0.0010)	(0.0010)	(0.0009)	(0.0011)	(0.0009)	(0.0009)
Gender	0.001	0.026	-0.043*	0.123***	0.164***	0.188***
	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)
Child's age	0.225***	0.159***	0.0947***	0.165***	0.108***	0.046***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)
Weight	0.072***	0.075***	0.063***	0.039***	0.055***	0.056***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)
Young mother	-0.096*	-0.134**	-0.127*	-0.082*	-0.067	-0.107
	(0.049)	(0.052)	(0.058)	(0.039)	(0.049)	(0.059)
Mature mother	-0.015	0.003	-0.009	-0.001	-0.012	-0.016
	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)
WIC	-0.127***	-0.125***	-0.132***	-0.133***	-0.107***	-0.131***
	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.023)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.378	0.360	0.367	0.338	0.330	0.359

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables and school fixed effects are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.7: Regression results – controls and school inputs included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.152***	-0.229***	-0.347***	-0.034	-0.067*	-0.109***
	(0.029)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)
Hispanic	-0.209***	-0.159***	-0.224***	-0.120***	-0.124***	-0.152***
	(0.025)	(0.026)	(0.027)	(0.025)	(0.025)	(0.027)
Asian	0.305***	0.189***	0.161***	0.366***	0.268***	0.179***
	(0.038)	(0.037)	(0.034)	(0.046)	(0.041)	(0.034)
Race else	-0.177*	-0.055	-0.118	-0.178**	-0.111	-0.063
	(0.069)	(0.070)	(0.080)	(0.061)	(0.064)	(0.083)
SES	0.293***	0.272***	0.234***	0.304***	0.278***	0.249***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)
Books	0.106***	0.097***	0.107***	0.106***	0.092***	0.077***
	(0.016)	(0.015)	(0.015)	(0.017)	(0.015)	(0.014)
Books^2	-0.0050***	-0.0045***	-0.0050***	-0.0050***	-0.0044***	-0.0036***
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0009)	(0.0010)
Gender	0.018	0.032	-0.022	0.132***	0.166***	0.212***
	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)
Child's age	0.250***	0.181***	0.112***	0.196***	0.124***	0.0647***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Weight	0.076***	0.079***	0.067***	0.042***	0.055***	0.059***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)
Young mother	-0.097*	-0.074	-0.080	-0.078	-0.028	-0.099
	(0.047)	(0.052)	(0.055)	(0.040)	(0.049)	(0.055)
Mature mother	0.006	0.003	-0.010	0.002	-0.029	-0.023
	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
WIC	-0.138***	-0.116***	-0.146***	-0.141***	-0.109***	-0.160***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Free lunch	-0.104***	-0.045***	-0.038**	-0.063***	-0.017	-0.037**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Public school	-0.013	-0.071*	0.014	0.030	0.012	0.013
	(0.030)	(0.029)	(0.028)	(0.033)	(0.032)	(0.029)
	⋮	⋮	⋮	⋮	⋮	⋮

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Size of class	0.011	0.014	0.024**	-0.007	0.002	0.024**
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Ratio of whites	-0.053***	-0.003	0.018	-0.078***	0.006	0.014
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ratio of blacks	0.000	0.004	0.031*	0.025*	0.032**	0.039**
	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.013)
Turnover$_1$	-0.084	-0.668***	-0.184	0.156	-0.297	-0.373
	(0.166)	(0.200)	(0.185)	(0.115)	(0.184)	(0.225)
Turnover$_2$	0.210	-0.472*	-0.0276	0.386***	-0.195	-0.208
	(0.157)	(0.190)	(0.174)	(0.108)	(0.176)	(0.214)
Turnover$_3$	0.204	-0.445*	-0.050	0.353***	-0.185	-0.227
	(0.156)	(0.189)	(0.174)	(0.106)	(0.175)	(0.214)
Security	0.019	0.037	-0.017	-0.016	0.003	-0.030
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)
Internet access	-0.0693	-0.121	-0.152	0.0476	-0.0505	-0.166
	(0.138)	(0.108)	(0.136)	(0.137)	(0.147)	(0.132)
Computers	0.022**	0.030***	0.024**	0.032***	0.049***	0.036***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Drugs$_1$	0.059	0.188***	0.104*	0.052	0.194***	0.148**
	(0.045)	(0.053)	(0.053)	(0.042)	(0.046)	(0.053)
Drugs$_2$	0.107*	0.210***	0.203***	0.113*	0.234***	0.217***
	(0.049)	(0.056)	(0.056)	(0.046)	(0.049)	(0.056)
Gang problems$_1$	0.046	-0.025	0.021	0.033	-0.093	0.055
	(0.055)	(0.062)	(0.063)	(0.049)	(0.055)	(0.064)
Gang problems$_2$	0.057	0.000	0.013	0.056	-0.090	0.063
	(0.060)	(0.066)	(0.068)	(0.055)	(0.061)	(0.068)
Constant	-0.161	0.523*	0.187	-0.544**	0.105	0.161
	(0.215)	(0.224)	(0.227)	(0.179)	(0.234)	(0.257)
N	13588	13588	13588	13588	13588	13588
R^2	0.291	0.235	0.211	0.216	0.180	0.178

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables and school inputs are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.8: Regression results – controls, school and family inputs included

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.152***	-0.227***	-0.343***	-0.036	-0.066*	-0.107***
	(0.029)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)
Hispanic	-0.194***	-0.143***	-0.204***	-0.103***	-0.101***	-0.128***
	(0.025)	(0.026)	(0.027)	(0.025)	(0.025)	(0.027)
Asian	0.313***	0.198***	0.173***	0.382***	0.286***	0.194***
	(0.038)	(0.037)	(0.034)	(0.046)	(0.040)	(0.035)
Race else	-0.182**	-0.060	-0.121	-0.183**	-0.121	-0.069
	(0.070)	(0.071)	(0.080)	(0.061)	(0.064)	(0.082)
SES	0.279***	0.257***	0.221***	0.293***	0.262***	0.232***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)
Books	0.095***	0.085***	0.095***	0.089***	0.070***	0.060***
	(0.015)	(0.015)	(0.015)	(0.016)	(0.015)	(0.014)
Books^2	-0.0045***	-0.0039***	-0.0045***	-0.0044***	-0.0035***	-0.0028**
	(0.0011)	(0.0010)	(0.0010)	(0.0011)	(0.0008)	(0.0009)
Gender	0.017	0.030	-0.025	0.128***	0.161***	0.209***
	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)
Child's age	0.250***	0.182***	0.113***	0.197***	0.126***	0.066***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Weight	0.076***	0.079***	0.067***	0.042***	0.056***	0.059***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)
Young mother	-0.090	-0.065	-0.064	-0.081*	-0.024	-0.087
	(0.047)	(0.051)	(0.055)	(0.041)	(0.049)	(0.055)
Mature mother	0.010	0.006	-0.007	0.004	-0.027	-0.019
	(0.018)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
WIC	-0.130***	-0.108***	-0.135***	-0.133***	-0.100***	-0.145***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
	⋮	⋮	⋮	⋮	⋮	⋮

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Free Lunch	-0.099***	-0.039**	-0.031*	-0.058***	-0.010	-0.029*
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Public school	-0.020	-0.078**	0.006	0.028	0.003	0.002
	(0.030)	(0.029)	(0.028)	(0.033)	(0.032)	(0.029)
Size of class	0.009	0.013	0.023**	-0.008	0.001	0.023**
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Ratio of whites	-0.052***	-0.004	0.017	-0.076***	0.007	0.014
	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Ratio of blacks	0.001	0.005	0.032**	0.025*	0.033**	0.040**
	(0.011)	(0.012)	(0.012)	(0.011)	(0.012)	(0.013)
Turnover$_1$	-0.095	-0.662***	-0.182	0.149	-0.296	-0.377
	(0.164)	(0.195)	(0.180)	(0.114)	(0.180)	(0.220)
Turnover$_2$	0.198	-0.478*	-0.0297	0.379***	-0.195	-0.214
	(0.156)	(0.186)	(0.169)	(0.107)	(0.173)	(0.210)
Turnover$_3$	0.196	-0.448*	-0.048	0.349***	-0.182	-0.228
	(0.155)	(0.185)	(0.168)	(0.105)	(0.172)	(0.209)
Security	0.021	0.038	-0.014	-0.013	0.006	-0.027
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)
Internet access	-0.061	-0.113	-0.152	0.053	-0.043	-0.161
	(0.136)	(0.107)	(0.136)	(0.135)	(0.149)	(0.133)
Computers	0.020*	0.028**	0.022*	0.031***	0.048***	0.034***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Drugs$_1$	0.057	0.185***	0.097	0.049	0.188***	0.143**
	(0.044)	(0.052)	(0.052)	(0.042)	(0.046)	(0.051)
Drugs$_2$	0.104*	0.205***	0.194***	0.110*	0.228***	0.209***
	(0.048)	(0.055)	(0.055)	(0.046)	(0.049)	(0.054)
Gang problems$_1$	0.037	-0.032	0.016	0.026	-0.101	0.046
	(0.055)	(0.061)	(0.062)	(0.049)	(0.055)	(0.062)
Gang problems$_2$	0.046	-0.010	0.004	0.043	-0.103	0.048
	(0.060)	(0.066)	(0.067)	(0.055)	(0.060)	(0.067)
	⋮	⋮	⋮	⋮	⋮	⋮

	Mathematics			Reading		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Reading time	0.002	-0.004	-0.019*	0.015	0.013	-0.010
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)
Tell stories$_1$	0.056	0.011	-0.008	-0.000	-0.044	0.037
	(0.053)	(0.060)	(0.059)	(0.062)	(0.062)	(0.061)
Tell stories$_2$	0.038	0.021	-0.015	-0.030	-0.043	0.008
	(0.053)	(0.060)	(0.059)	(0.063)	(0.062)	(0.060)
Tell stories$_3$	0.046	0.031	-0.014	0.008	-0.005	0.051
	(0.054)	(0.061)	(0.060)	(0.064)	(0.063)	(0.061)
Book reading$_1$	0.011	0.099	0.145	-0.007	0.088	0.164
	(0.084)	(0.095)	(0.102)	(0.083)	(0.080)	(0.103)
Book reading$_2$	0.083	0.165	0.265**	0.049	0.190*	0.272**
	(0.082)	(0.094)	(0.100)	(0.082)	(0.079)	(0.101)
Book reading$_3$	0.126	0.206*	0.308**	0.135	0.265***	0.334***
	(0.083)	(0.094)	(0.100)	(0.083)	(0.079)	(0.101)
Computer home	0.138***	0.132***	0.124***	0.084***	0.116***	0.156***
	(0.021)	(0.022)	(0.023)	(0.021)	(0.020)	(0.023)
Constant	-0.395	0.230	-0.155	-0.673***	-0.164	-0.266
	(0.230)	(0.241)	(0.245)	(0.197)	(0.247)	(0.275)
N	13588	13588	13588	13588	13588	13588
R^2	0.295	0.239	0.217	0.220	0.188	0.186

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables, school and family inputs are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.9: Regression results – controls, additional inputs with school fixed effects

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.158***	-0.235***	-0.308***	-0.048	-0.078*	-0.079*
	(0.031)	(0.034)	(0.036)	(0.033)	(0.033)	(0.035)
Hispanic	-0.213***	-0.166***	-0.176***	-0.139***	-0.118***	-0.116***
	(0.026)	(0.028)	(0.030)	(0.026)	(0.027)	(0.030)
Asian	0.269***	0.194***	0.128**	0.373***	0.307***	0.211***
	(0.044)	(0.044)	(0.041)	(0.050)	(0.046)	(0.040)
Race Else	-0.131	-0.061	-0.088	-0.069	-0.086	-0.062
	(0.078)	(0.081)	(0.089)	(0.070)	(0.069)	(0.092)
SES	0.245***	0.234***	0.204***	0.256***	0.247***	0.217***
	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)
Books	0.087***	0.077***	0.086***	0.093***	0.066***	0.047**
	(0.015)	(0.015)	(0.015)	(0.016)	(0.015)	(0.014)
Books^2	-0.0043***	-0.0037***	-0.0041***	-0.0045***	-0.0031***	-0.0021*
	(0.0009)	(0.0010)	(0.0009)	(0.0010)	(0.0008)	(0.0009)
Gender	0.008	0.025	-0.045*	0.119***	0.160***	0.186***
	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)
Child's age	0.226***	0.160***	0.095***	0.166***	0.109***	0.045***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)
Weight	0.072***	0.076***	0.063***	0.040***	0.056***	0.056***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)
Young mother	-0.090	-0.127*	-0.115*	-0.087*	-0.064	-0.098
	(0.048)	(0.051)	(0.058)	(0.040)	(0.049)	(0.059)
Mature mother	-0.014	0.003	-0.004	0.002	-0.012	-0.010
	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)
WIC	-0.119***	-0.117***	-0.131***	-0.123***	-0.099***	-0.125***
	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.023)
	⋮	⋮	⋮	⋮	⋮	⋮

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Free lunch	0.041	0.022	0.070**	-0.008	-0.029	0.059**
	(0.044)	(0.268)	(0.022)	(0.038)	(0.230)	(0.018)
Public school	0.294	-0.124	-0.028	-0.082	-0.026	0.003
	(0.370)	(0.248)	(0.045)	(0.511)	(0.251)	(0.040)
Size of class	0.046*	0.068***	0.026	0.056**	0.082***	0.029*
	(0.019)	(0.020)	(0.013)	(0.019)	(0.019)	(0.013)
Ratio of whites	0.032	-0.059	-0.089***	0.036	0.082	-0.012
	(0.047)	(0.095)	(0.025)	(0.043)	(0.098)	(0.023)
Ratio of blacks	0.036	-0.061	0.009	0.064	-0.029	0.003
	(0.034)	(0.083)	(0.018)	(0.034)	(0.083)	(0.016)
Turnover$_1$	0.931***	0.827	0.155	0.116	2.021*	0.176
	(0.234)	(1.205)	(0.230)	(0.300)	(0.941)	(0.224)
Turnover$_2$	0.809***	2.300***	0.101	0.345	1.209*	0.074
	(0.183)	(0.557)	(0.223)	(0.253)	(0.494)	(0.209)
Turnover$_3$	0.894***	1.740***	0.111	0.459	1.869***	0.066
	(0.166)	(0.393)	(0.220)	(0.242)	(0.345)	(0.206)
Security	-0.054	-0.098	-0.032	-0.026	-0.043	-0.012
	(0.056)	(0.067)	(0.044)	(0.057)	(0.068)	(0.046)
Internet access	0.441	0.914*	0.239	0.067	0.007	0.053
	(0.226)	(0.447)	(0.205)	(0.166)	(0.616)	(0.150)
Computers	-0.023	0.010	0.001	0.026	-0.084	0.015
	(0.029)	(0.089)	(0.018)	(0.027)	(0.081)	(0.018)
Drugs$_1$	-0.186	0.650	0.032	-0.152	0.422	0.053
	(0.132)	(1.255)	(0.0743)	(0.132)	(0.995)	(0.067)
Drugs$_2$	-0.096	0.701	0.043	-0.111	0.702	0.158*
	(0.151)	(1.226)	(0.083)	(0.140)	(0.973)	(0.076)
Gang problems$_1$	0.174	-0.143	0.098	0.005	-0.014	0.045
	(0.156)	(0.525)	(0.106)	(0.119)	(0.519)	(0.093)
Gang problems$_2$	0.005	-0.404	0.107	-0.138	-0.384	-0.047
	(0.182)	(1.133)	(0.116)	(0.137)	(0.893)	(0.102)
	⋮	⋮	⋮	⋮	⋮	⋮

	Mathematics			Reading		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Reading time	0.005	0.001	-0.015	0.017*	0.019*	-0.008
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
Tell stories$_1$	0.025	-0.028	-0.039	-0.049	-0.071	0.021
	(0.054)	(0.062)	(0.062)	(0.061)	(0.059)	(0.062)
Tell stories$_2$	0.025	0.003	-0.036	-0.063	-0.057	-0.010
	(0.055)	(0.062)	(0.062)	(0.062)	(0.060)	(0.062)
Tell stories$_3$	0.034	0.008	-0.037	-0.018	-0.026	0.028
	(0.056)	(0.063)	(0.062)	(0.063)	(0.061)	(0.062)
Book reading$_1$	0.038	0.062	0.130	0.007	0.100	0.188
	(0.082)	(0.096)	(0.116)	(0.078)	(0.083)	(0.117)
Book reading$_2$	0.082	0.098	0.224*	0.040	0.172*	0.279*
	(0.081)	(0.095)	(0.114)	(0.077)	(0.082)	(0.115)
Book reading$_3$	0.126	0.140	0.266*	0.134	0.239**	0.331**
	(0.082)	(0.095)	(0.114)	(0.078)	(0.082)	(0.115)
Computer home	0.140***	0.123***	0.121***	0.094***	0.115***	0.152***
	(0.021)	(0.022)	(0.024)	(0.021)	(0.021)	(0.024)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	13588	13588	13588	13588	13588	13588
R^2	0.383	0.365	0.373	0.344	0.337	0.366

Notes: The columns report weighted OLS regression results. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The standardized baseline control variables, school and family inputs, as well as school fixed effects are used as regressors. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.10: Regression results – robustness checks

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	0.119***	0.282***	0.218***	0.193***	0.176***	0.245***
	(0.032)	(0.026)	(0.027)	(0.029)	(0.031)	(0.031)
Black	-0.319***	-0.359***	-0.279***	-0.374***	-0.337***	-0.361***
	(0.036)	(0.040)	(0.079)	(0.032)	(0.036)	(0.040)
Hispanic	-0.269***	-0.225***	-0.271***	-0.270***	-0.282***	-0.258***
	(0.032)	(0.038)	(0.033)	(0.034)	(0.035)	(0.033)
Asian	0.187***	0.139***	0.158***	0.151***	0.176***	0.123*
	(0.055)	(0.040)	(0.046)	(0.045)	(0.042)	(0.049)
Race Else	-0.230*	-0.073	-0.088	-0.399**	-0.116	-0.230
	(0.092)	(0.136)	(0.0953)	(0.121)	(0.0933)	(0.128)
SES	0.210***	0.227***	0.262***	0.237***	0.247***	0.253***
	(0.021)	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)
Books	0.197***	0.079***	0.113***	0.118***	0.103***	0.114***
	(0.028)	(0.016)	(0.017)	(0.025)	(0.021)	(0.021)
Books^2	-0.0191***	-0.0035***	-0.0048***	-0.0076**	-0.0045***	-0.0054***
	(0.0036)	(0.0008)	(0.0010)	(0.0028)	(0.0013)	(0.0014)
Gender	0.022	-0.070**	-0.034	-0.008	-0.007	-0.039
	(0.026)	(0.023)	(0.024)	(0.025)	(0.025)	(0.024)
Child's age	0.089***	0.143***	0.117***	0.111***	0.107***	0.120***
	(0.014)	(0.012)	(0.014)	(0.013)	(0.013)	(0.014)
Weight	0.070***	0.062***	0.076***	0.058***	0.070***	0.056**
	(0.013)	(0.012)	(0.012)	(0.012)	(0.016)	(0.020)
Young mother	-0.069	0.003	-0.087	-0.081	-0.040	-0.132
	(0.060)	(0.129)	(0.085)	(0.073)	(0.070)	(0.089)
Mature mother	-0.006	-0.023	-0.022	0.012	-0.001	-0.013
	(0.030)	(0.024)	(0.026)	(0.027)	(0.027)	(0.027)
WIC	-0.129***	-0.142***	-0.187***	-0.137***	-0.128***	-0.200***
	(0.030)	(0.030)	(0.029)	(0.029)	(0.030)	(0.028)
N	6843	6745	7113	6475	6806	6782
R^2	0.110	0.146	0.200	0.196	0.178	0.221

Notes: The columns report weighted OLS regression results. The standardized mathematics test score in the spring of 2011 is the dependent variable. The racial attribute dummies and standardized baseline control variables are used as regressors. Robustness checks (1) and (2) are regarding children who have below and above the median socio-economic status values. Specifications (3) and (4) are separated based on the below and above the median ratio of black students in the school. Models (5) and (6) are estimated on the subsample of below and above median weight children. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.11: Regression results – black race specific models

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	-10.287**	8.450	38.643***	10.021**	20.967***	54.463***
	(3.378)	(4.496)	(5.052)	(3.136)	(4.289)	(5.410)
SES	3.946***	4.402***	4.443***	3.592***	4.231***	4.901***
	(0.337)	(0.384)	(0.499)	(0.302)	(0.401)	(0.510)
Books/1000	20.626***	11.384	6.549	13.512**	22.448**	9.191
	(5.544)	(6.785)	(7.479)	(4.970)	(7.447)	(7.609)
(Books/1000)^2	-29.870**	-10.571	-7.687	-15.803*	-28.774**	-13.137
	(9.149)	(13.20)	(11.08)	(6.530)	(8.891)	(8.648)
Gender	0.357	0.628	0.364	0.983**	2.176***	2.962***
	(0.417)	(0.500)	(0.551)	(0.377)	(0.508)	(0.571)
Child's age	0.542***	0.438***	0.291***	0.393***	0.385***	0.188*
	(0.045)	(0.060)	(0.070)	(0.039)	(0.056)	(0.074)
Weight	0.023*	0.026	0.016	0.008	0.012	0.012
	(0.011)	(0.013)	(0.014)	(0.008)	(0.012)	(0.015)
Young mother	-0.449	0.223	-2.537	-0.241	-0.481	-2.334
	(0.867)	(0.981)	(1.359)	(0.643)	(0.939)	(1.213)
Mature mother	0.446	0.534	-0.160	0.195	0.122	-0.142
	(0.478)	(0.563)	(0.622)	(0.426)	(0.569)	(0.626)
WIC	-2.001***	-1.657**	-1.815**	-1.689**	-1.722*	-2.456***
	(0.541)	(0.623)	(0.649)	(0.550)	(0.684)	(0.686)
N	2038	2038	2038	2038	2038	2038
R^2	0.186	0.132	0.089	0.163	0.129	0.110

Notes: The columns report weighted OLS regression results. Mathematics and reading unstandardized test scores of black students in different time periods are the dependent variables. The baseline unstandardized control variables are used as regressors. The school clustered standard errors are included in the parentheses. Stars denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.12: Regression results – white race specific models

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Constant	-18.782***	2.137	31.565***	4.438**	19.690***	48.364***
	(1.916)	(2.186)	(2.433)	(1.582)	(1.988)	(2.317)
SES	4.668***	4.499***	4.977***	3.601***	4.303***	5.021***
	(0.166)	(0.178)	(0.216)	(0.149)	(0.187)	(0.215)
Books/1000	11.708***	11.735***	14.807***	9.746***	11.362***	11.406***
	(1.668)	(1.735)	(2.278)	(1.502)	(1.730)	(2.070)
$(\text{Books}/1000)^2$	-4.328***	-4.209***	-5.622***	-3.591***	-4.084***	-4.079***
	(0.969)	(0.957)	(1.156)	(0.770)	(0.794)	(1.003)
Gender	0.220	0.393	-0.320	1.255***	1.878***	2.543***
	(0.204)	(0.221)	(0.250)	(0.187)	(0.228)	(0.250)
Child's age	0.650***	0.522***	0.385***	0.434***	0.360***	0.221***
	(0.026)	(0.029)	(0.032)	(0.021)	(0.026)	(0.030)
Weight	0.047***	0.053***	0.054***	0.025***	0.041***	0.049***
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
Young mother	-1.250*	-1.206	-0.624	-0.877	-0.356	-0.916
	(0.603)	(0.728)	(0.813)	(0.460)	(0.725)	(0.831)
Mature mother	0.117	-0.00447	-0.198	0.116	-0.438	-0.373
	(0.222)	(0.241)	(0.272)	(0.204)	(0.250)	(0.268)
WIC	-1.916***	-1.771***	-2.499***	-1.378***	-1.387***	-2.416***
	(0.251)	(0.268)	(0.293)	(0.217)	(0.275)	(0.289)
N	10184	10184	10184	10184	10184	10184
R^2	0.270	0.215	0.181	0.200	0.166	0.164

Notes: The columns report weighted OLS regression results. Mathematics and reading unstandardized test scores of white students in different time periods are the dependent variables. The baseline unstandardized control variables are used as regressors. The school clustered standard errors are included in the parentheses. Stars denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.13: Regression results – score gap after propensity score matching

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
Black	-0.032	-0.145*	-0.319***	0.073	-0.030	-0.085
	(0.053)	(0.061)	(0.054)	(0.046)	(0.057)	(0.056)
SES	0.222***	0.245***	0.187***	0.225***	0.314***	0.265***
	(0.036)	(0.036)	(0.033)	(0.033)	(0.042)	(0.037)
Books	0.218***	0.150*	0.206**	0.188**	0.097	0.096
	(0.061)	(0.074)	(0.076)	(0.059)	(0.062)	(0.076)
Books^2	-0.045***	-0.026	-0.036*	-0.029*	-0.020	-0.018
	(0.014)	(0.021)	(0.017)	(0.012)	(0.014)	(0.015)
Gender	-0.015	0.022	0.028	0.076	0.114	0.185**
	(0.053)	(0.061)	(0.058)	(0.048)	(0.060)	(0.062)
Child's age	0.213***	0.147***	0.010	0.153***	0.079*	0.002
	(0.035)	(0.039)	(0.056)	(0.026)	(0.036)	(0.050)
Weight	0.051	0.046	0.023	0.037	0.042	0.045
	(0.027)	(0.028)	(0.042)	(0.021)	(0.026)	(0.038)
Young mother	-0.502***	-0.430**	0.098	-0.138*	-0.280*	-0.087
	(0.133)	(0.166)	(0.098)	(0.058)	(0.122)	(0.083)
Mature mother	-0.040	-0.027	-0.155*	0.040	-0.056	-0.121
	(0.057)	(0.067)	(0.066)	(0.059)	(0.067)	(0.075)
WIC	-0.082	-0.082	-0.091	-0.047	-0.039	-0.068
	(0.049)	(0.050)	(0.051)	(0.047)	(0.053)	(0.053)
Free lunch	-0.107***	-0.050	-0.059*	-0.083***	-0.059	-0.081*
	(0.027)	(0.030)	(0.030)	(0.025)	(0.032)	(0.033)
Public school	-0.104	-0.112	-0.032	0.018	0.091	-0.010
	(0.077)	(0.077)	(0.075)	(0.074)	(0.084)	(0.082)
Size of class	0.003	-0.009	0.032	-0.060**	-0.025	0.001
	(0.025)	(0.030)	(0.027)	(0.023)	(0.023)	(0.028)
	⋮	⋮	⋮	⋮	⋮	⋮

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Ratio of whites	0.023	0.022	-0.011	0.017	0.029	0.017
	(0.041)	(0.045)	(0.053)	(0.037)	(0.043)	(0.056)
Ratio of blacks	0.036	0.019	0.025	0.059*	0.028	0.063*
	(0.026)	(0.029)	(0.029)	(0.027)	(0.032)	(0.032)
Turnover$_1$	-0.281	-0.618*	-0.222	0.609**	-0.276	-0.613
	(0.317)	(0.302)	(0.245)	(0.194)	(0.249)	(0.359)
Turnover$_2$	0.0792	-0.364	-0.0713	0.836***	-0.104	-0.626
	(0.292)	(0.264)	(0.209)	(0.180)	(0.232)	(0.326)
Turnover$_3$	-0.0172	-0.381	-0.132	0.749***	-0.102	-0.565
	(0.295)	(0.269)	(0.215)	(0.178)	(0.230)	(0.330)
Security	0.078	0.096	-0.061	-0.070	-0.068	-0.144*
	(0.061)	(0.071)	(0.070)	(0.058)	(0.072)	(0.071)
Internet access	-0.636	-0.338	-0.156	0.067	0.479*	-0.0422
	(0.621)	(0.274)	(0.375)	(0.248)	(0.230)	(0.305)
Computers	-0.012	-0.017	-0.004	-0.059	-0.041	-0.019
	(0.019)	(0.025)	(0.020)	(0.033)	(0.035)	(0.036)
Drugs$_1$	0.011	-0.001	-0.043	0.102	-0.013	-0.016
	(0.136)	(0.148)	(0.104)	(0.110)	(0.119)	(0.104)
Drugs$_2$	0.093	0.079	0.047	0.247*	0.116	0.063
	(0.144)	(0.156)	(0.109)	(0.119)	(0.130)	(0.111)
Gang problems$_1$	-0.022	-0.111	-0.014	-0.066	-0.134	0.097
	(0.143)	(0.158)	(0.140)	(0.121)	(0.147)	(0.144)
Gang problems$_2$	-0.081	-0.100	0.055	-0.200	-0.206	0.118
	(0.150)	(0.164)	(0.147)	(0.139)	(0.163)	(0.159)
	⋮	⋮	⋮	⋮	⋮	⋮

	MATHEMATICS			READING		
	2010 Fall	2011 Spring	2012 Spring	2010 Fall	2011 Spring	2012 Spring
	⋮	⋮	⋮	⋮	⋮	⋮
Reading time	-0.004	-0.024	0.002	0.022	0.005	0.006
	(0.018)	(0.022)	(0.020)	(0.016)	(0.018)	(0.020)
Tell stories$_1$	-0.138	0.015	0.066	-0.242	-0.252	0.077
	(0.137)	(0.128)	(0.123)	(0.142)	(0.146)	(0.123)
Tell stories$_2$	-0.108	-0.006	0.050	-0.260	-0.256	0.012
	(0.140)	(0.129)	(0.125)	(0.146)	(0.150)	(0.126)
Tell stories$_3$	-0.273	-0.032	-0.166	-0.310*	-0.276	-0.106
	(0.144)	(0.133)	(0.133)	(0.147)	(0.151)	(0.130)
Book reading$_1$	-0.008	-0.087	-0.068	-0.082	0.077	0.014
	(0.226)	(0.279)	(0.225)	(0.276)	(0.243)	(0.183)
Book reading$_2$	0.197	0.136	0.018	0.007	0.301	0.154
	(0.224)	(0.274)	(0.223)	(0.277)	(0.242)	(0.179)
Book reading$_3$	0.296	0.320	0.220	0.103	0.449	0.297
	(0.221)	(0.271)	(0.214)	(0.277)	(0.240)	(0.172)
Computer home	0.083	0.019	0.021	0.100*	-0.057	0.059
	(0.061)	(0.069)	(0.068)	(0.047)	(0.069)	(0.069)
Constant	0.651	0.696	0.313	-0.789	-0.340	0.281
	(0.737)	(0.490)	(0.522)	(0.419)	(0.425)	(0.518)
N	4076	4076	4076	4076	4076	4076
R^2	0.247	0.183	0.114	0.167	0.156	0.120

Notes: Columns report OLS regression results after propensity score matching. Mathematics and reading standardized test scores of students in different time periods are the dependent variables. The racial attributes, controls, family and school inputs are used as regressors, while multiracial students are excluded from the sample. The school clustered standard errors are included in the parentheses. Asterisks denote empirical significance levels according to: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.