KATHOLIEKE UNIVERSITEIT LEUVEN

How schools influence students' academic achievements: a behavioral approach with empirical evidence from add health data

by

Yuemei JI

Public Economics

DISCUSSION PAPER

How schools influence students' academic achievements

A Behavioral Approach with Empirical Evidence from Add Health Data

Yuemei JI
yuemei.ji@econ.kuleuven.be
Center for Economic Studies, KULeuven

Abstract

This paper proposes a behavioral model to study how schools influence students' educational behavior and academic achievements. The school quality is then defined into two dimensions: the amount of market-valued skills schools impart and how well schools cultivate an educational identity. Using data from Add Health in the US, I test the major hypotheses from the theoretical model. On the one hand, school resources (average class size and teacher supply) and student-level curriculum have some effects on the math GPA scores. On the other hand, educational identity indicators (school-level happiness and participation at school teams, clubs or organizations) and the previous math GPA scores are significant determinants in students' observable effort level such as absenteeism behavior, and through this channel both determinants indirectly influence math GPA achievement. These empirical results inform us that an identity-based behavioral model adds to a rational expectation educational choice model in understanding the widening academic achievement gap between adolescents from different socioeconomic backgrounds. The paper presents the limitation of using school resources to study the school quality and advocates a richer set of school quality measures.

Part I   Introduction

Since the end of the 1970s, overall wage inequality and educational differentials have expanded in most OECD countries. This trend is coupled with a rise in the payoff to high education and skills as the demand for high quality labor exceeds the inadequate supply. Despite substantial increasing premiums to high education, the college participation rates in the US increase more sharply in the high income groups than in the low ones (Atkinson, 2003). The puzzle is why, if to improve education and skills becomes more financially rewarding, there are still high drop-out rates at high school among the economically disadvantaged youth.

One plausible explanation to this puzzle is the ever increasing social segregation and the important role it plays in the deterioration of schools in poor neighborhoods (Benabou, 1993). Burtless (1996) and Kozol (1991) point out that children living in a poor segregated community can only enter low quality schools with insufficient school expenditure on facilities, classes and teachers. A number of empirical studies have investigated the impact of school resources (characteristics of schools, school expenditures, class size, qualifications of teachers) on students' academic achievement (Hanushek et al. (2006)).

Two criticisms have recently been leveled with respect to the theoretical education production function underlying the current empirical studies on school resources and performance. The first one is that these school resources studies all test that the output of the educational process is closely related to school inputs. However, the added resources to schools in many countries actually can be ineffective as the market-valued cognitive skills such as mathematics and reading skills which one obtains from these school resources can be very low. Using micro datasets, the school resources literature has put a lot of efforts at studying the effects of class size, per pupil expenditure, teacher education and experience on the improvement of educational performance. Hanushek (2003) concludes that many of these empirical studies only focus on a small set of direct school input measures but empirically have not identified any noticeable impacts that lead to understanding the significant gaps between the advantaged and disadvantaged schools in the US. At the aggregate level, Hanushek (2003) reveals that real per student expenditure doubles

from 1970 to 2000 in the US, and by contrast, the average performance of students who are 17 years old in mathematics and reading is only slightly higher in 1999 than 30 years before.

The second cutting-edge problem is that these school resources studies do not pay sufficient attention to students' internal educational motivation. It is often implicitly assumed that individuals have a rational expectation of the future returns to educational investment at school. When a good education is economically rewarding in the labor market, students should positively react to this external incentive by making enough effort at the school. This assumption has not yet been empirically supported as future income expectations are heterogeneous in nature (Dominitz and Manski (1996)). A number of empirical studies have shown that preschool education, family socioeconomic status, peer effects and non-cognitive ability greatly affect one's educational performances (Hanushek et al. (2006) & Heckman et al. (2006)). It is suggested that these factors probably do not directly influence learning but are influenced by the school norms that affect motivations and behaviors of students. Akerlof and Kranton (2002) and Ji (2008) both argue that emotion such as self-esteem driven by an educational identity greatly influences judgment and it acts as a motivational input into the schooling decision making process.

Following the same approach, this paper firstly proposes a model which allows for a discussion of the relationship between school effects and academic achievements. School organization does not act as a sheer place to impart cognitive skills but has its social settings and strategy to cultivate an educational identity through the interactions among school, individuals and other peers. The hypothesis is that there are three determinants of individual effort and academic achievement. They are the amount of market-valued cognitive skills which the school imparts, the intensity (salience) of educational identity the school promotes and previous academic achievement. This behavioral educational choice model offers us some broader insights into the hidden problems of American schools. In particular, the educational identity dimension allows us to explain questions such as why peer effects and school social composition matter. To test these theoretical findings, I use a dataset from the National Longitudinal Study of Adolescent Health. It provides detailed and comparable information such as current and previous math GPA scores,

student-level math curriculum, school social composition, feeling of happiness and participation at school activities.

This study distinguishes itself from other school studies in the following sense. It is the first to study the effects of schools based on a behavioral education production model. Instead of assuming schooling production process as a black-box[1], it explores the hidden mechanisms and adds an important dimension—educational identity—to the empirical study of school quality. By comparing a set of different school quality measures, it shows the limitation of using school resources to determine the school quality and advocates a set of measures of school quality (student-level curriculum content, school-level happiness and participation at school teams, clubs or organizations).

The organization of the paper is as follows. The following part presents a theoretical model and its major hypotheses. In Part III, the paper introduces the sample data, variables and the empirical strategy. Part IV discusses the empirical results and it finally concludes in Part V.

Part II   Theoretical Model

Schools are specialized organizations where students invest in their human capital. According to the standard human capital theory (Becker (1964)), individuals make choices of investing in human capital based on a rational benefits and costs analysis. When the returns to an investment exceed the costs induced, the individual will choose to invest in human capital. The behavioral model in this paper introduces a subjective utility term to the educational utility function. It assumes that the individuals not only considers how much they can earn in the future labor market with the human capital they accumulate at school, but also other non-pecuniary concerns such as self-esteem linked to academic achievements[2]. To be specific, the utility function for student i at school j writes as,

---

[1] There is a large set of studies in the economics of education literature. These studies have identified considerable impacts of peer effects and social interactions on students' academic achievement. But most of them such as Hoxby(2000)   treat schooling process as a black box without clarifying in which mechanism peers at school influence each other. From this perspective, the links between theoretical development and empirical work are not strong. Peers effects and interactions can probably work through two channels: (1) direct learning and knowledge sharing among students; (2) externality of motivation. This paper discussion the second channel of peer effects.

[2] The utility assumption in this paper is less strong. I don't assume individuals have elaborate calculation of their future long term income from schooling. The utility presents here is a type of "cognitive-evaluative" view.

$$U_{ij} = IQ_i S_{ij} + q_j v(e_{ij}) S_{ij} - c(IQ_i, e_{ij}) \qquad (1)$$

$$S_{ij} = S_{ij}^0 + e_{ij} r_j \qquad (2)$$

The student's utility is composed of three parts as shown in the utility function (1). Firstly, the student's utility is directly determined by future income in the labor market. $IQ_i S_{ij}$ is the wage the student expects to receive in the future labor market. These economic returns equal the student's productivity level which is only determined by market-valued skills S_{ij} the student accumulates from school and her innate cognitive ability IQ_i. The future market uncertainties and risks are not taken into account[3]. The educational production of market-valued skills S_{ij} is described in function (2). r_j represents the quantity of cognitive skills the school *j* imparts to its student at a particular time period t. It is only associated with the school teaching content and quality then. For simplicity, assume that students learn knowledge and skills only from their teachers and they do not learn from each other. The higher r_j, the better the school is at teaching. For student *i* at school *j*, the accumulation of cognitive skills S_{ij} then equal the sum of the previous educational achievement S_{ij}^0 at time t-1 and the product of the student's effort input e_{ij} and the cognitive skills r_j. It is noticeable that the educational production process does not only depend on school but also on efforts exerted by the student to gain cognitive skills. Let $S_{ij}^0 > 0$ and $r_j > 0$.

Apart from the economic returns and costs described above in function (1), the student also takes into account self-esteem concerns $q_j v(e_{ij}) S_{ij}$ for engaging in schooling[4]. To be specific, $v(e_{ij})$ indicates how effort level e_{ij} determines the subjective way the student values the cognitive skills S_{ij}. She can then obtain from S_{ij} a subjective utility $v(e_{ij}) S_{ij}$. It is assumed that $v(.)$ increases in e_{ij}, $v(e_H) > v(e_L)$. The higher the effort the student exerts in schooling, the higher she values it. Furthermore, the self-esteem concerns at school are driven by the educational identity q_j.

---

[3] By contrast, in the standard human capital model, individual is assumed to have a rational expectation on future wage benefit from skills S_{ij}. The expected wage only based on one's cognitive ability IQ_i and labor market information on the wage distribution. The utility writes as $U_{ij} = Ew(IQ_i | \text{market information}) S_{ij} - c(IQ_i, e_{ij})$.

[4] In psychology, self-esteem reflects a person's overall evaluation of her own worth. William James (1890) used the term "self-esteem" to refer to the way individuals feel about themselves which in turn depends on the success that they wish to accomplish.

Assume that the educational identity q_j is a stable term in each school organization and students internalize this preference for schooling. Define q_j as the intensity (salience) of this educational identity for every student at school j. Some schools have a higher q_j than others. The higher q_j is, the greater the subjective utility for schooling S_{ij} is.

For simplicity, the student chooses between a high effort level e_H and a low effort level e_L. Denote $e_{ij} \in \{e_H, e_L\}$ and $e_H > e_L$. In equation (1), there is also an investment cost $c(\cdot)$. It is related to the student's cognitive ability IQ_i and effort input e_{ij}. Assume that $c(IQ_i, e_H) > c(IQ_i, e_L)$ and $\frac{\partial c(IQ_i, e_{ij})}{\partial IQ_i} < 0$.

Finally, to maximize the utility U_{ij} in utility function (1), student i chooses $e_{ij} = e_H$, when $U_{ij}(e_{ij} = e_H) \geq U_{ij}(e_{ij} = e_L)$; and $e_{ij} = e_L$ otherwise. The specific solutions to this simple maximization problem can be expressed as follows:

a) when $r_j \geq \frac{c(IQ_i, e_H) - c(IQ_i, e_L) - q_j S_{ij}^0 (v(e_H) - v(e_L))}{IQ_i (e_H - e_L) + q_j (v(e_H) e_H - v(e_L) e_L)}$, $(e_{ij}, v(e_{ij})) = (e_H, v(e_H))$;

b) when $r_j < \frac{c(IQ_i, e_H) - c(IQ_i, e_L) - q_j S_{ij}^0 (v(e_H) - v(e_L))}{IQ_i (e_H - e_L) + q_j (v(e_H) e_H - v(e_L) e_L)}$, $(e_{ij}, v(e_{ij})) = (e_L, v(e_L))$.

Obviously, the optimal effort level is influenced by (r_j, q_j, S_{ij}^0)[5]. The intuition shown in solutions a) and b) can be explained as follows. Given the level of the previous schooling S_{ij}^0, the more cognitive skills r_j the student can obtain at school, the greater external incentives will encourage her to make high level effort e_H. She will not only obtain extra economic rewards in the future labor market, but also higher self-esteem derived from a higher r_j. Similarly, given the level of the previous schooling S_{ij}^0, the higher the educational identity q_j, the student experiences higher self-esteem utility, and therefore is more internally motivated to choose high effort level e_H. The external and internal incentives r_j and q_j are mutually substitutable to achieve a high effort equilibrium.

---

[5] The rational expectation human capital model predicts that given the level of cognitive ability, the individual effort level e_{ij} is only influenced by wage distribution in the labor market and school effects r_j, but not by S_{ij}^0 and q_j.

The relationship between r_j and q_j, for a given level of S_{ij}^0 , is described graphically in Figure 1. The mathematical proof of Figure 1 can be found in the Appendix. I first set $S_{ij}^0 = S_1$, curve S_1 represents the boundary between the high effort equilibrium $e_{ij} = e_H$ and the low effort equilibrium $e_{ij} = e_L$. When $r_j = r^*$, $q_j = 0$. When $r_j = 0$, $q_j = q_j^1$. As r_j declines, the curve increases in q_j . All points above curve S_1 are combinations of r_j and q_j that produce a high effort equilibrium while all points below curve S_1 are combinations of r_j and q_j that produce a low effort equilibrium. The figure informs us that as r_j (or q_j) increases, point (r_j , q_j) is more likely to be above curve S_1. It is noticeable that as long as the acquired cognitive skills $r_j > r^*$, making high effort is optimal even when there is no internal motivation $q_j = 0$.

Figure 1. S_{ij}^0, q_j, r_j and Effort level
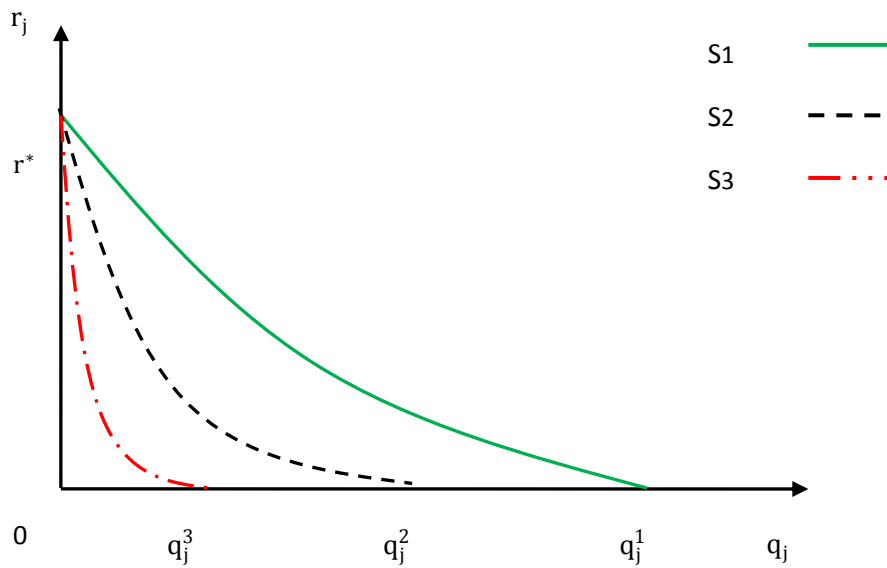
Figure 1 also shows that as initial skills S_{ij}^0 rise from S_1 to S_2, the boundary curve shifts downward from curve S_1 to curve S_2. When $r_j = r^*$, $q_j = 0$. When $r_j = 0$, $q_j = q_j^2$. As a result, the area of high effort equilibrium becomes larger and the area of low effort equilibrium shrinks. A further increase of initial skills S_{ij}^0 to S_3 will enlarge the area of high effort equilibrium further. Likewise, when $r_j = r^*$, $q_j = 0$. When $r_j = 0$, $q_j = q_j^3$. The intuition behind Figure 1 is an "endowment effect" mechanism. It has been assumed that $v(e_H) > v(e_L)$, the more efforts the

student makes in schooling, the higher value she will put on the total schooling $S_{ij}^0 + e_{ij} r_j$. For example, when the previous schooling has increased from S_1 to S_2, the differences in utility between making effort e_H and e_L are larger from S_2 than from S_1. As utility $q_j S_2 \left(v(e_H) - v(e_L) \right) > q_j S_1 \left(v(e_H) - v(e_L) \right)$, the high effort equilibrium condition in a) informs us that the student is more likely to make effort e_H when $S_{ij}^0 = S_2$ than when $S_{ij}^0 = S_1$.

Part III   Empirical Strategy

1. Data and Variables

The sample used for this study is from the National Longitudinal Study of Adolescent Health. Beginning at the 1994-95 school year, it is a longitudinal study of a nationally representative sample of adolescents in secondary school in the US. Within each participant school, the In-School Questionnaire was initially administered to all students who were present on the survey day. All students listed on the school roster are stratified by grade and sex, and randomly selected. Racial and ethnic minorities are oversampled and finally a total core sample of 12,105 adolescents was interviewed at home. Some students change or exit school after Wave I interviews. In the 1995-96 school year, Wave II does follow-up in-home interviews with the same adolescents who are still at their schools. These surveys provide a wide range of topics and information on social and demographic characteristics of respondents (such as gender, race and grade, education of parents, family composition), students' assessment on the relationships with their schools and teachers, self-reported academic behavior, performances and educational expectations at school. Other relevant datasets are also available for further and detailed studies. For example, administrators from participant schools complete self-administered questionnaires dealing with school policies and teacher characteristics. Moreover, in order to provide precise measures of math academic progress (GPA) and high school student-level curriculum, the Adolescent Health and Academic Achievement study collected school transcripts for some Wave I respondents throughout their grades 9-12.

The sample in this study is selected from students in grades 9-12 who participate in the Wave I In-home interview. Those with missing data on race, gender, IQ test, family and parents, school, student level math curriculum and math GPA scores are dropped from the sample. As the study is focusing on analyzing how school effects and previous math performance work on students' math GPA scores and absenteeism behavior, it requires students' detailed information on math curriculum and at least two consecutive math GPA scores during the academic years 1994-95 or 1995-96. As a result, a large set of observations are excluded. Table 1 presents the respondents' grade levels in the survey year 1994-95 and 1995-96 that were included in the sample. For a full list of variables, their definitions and sources in this study, refer to Table 2. The final sample is an unbalanced two-wave panel dataset with 5937 observations in total. Table 3 shows the sample structure: in Wave I there are 3066 observations; in Wave II 2871 and among them, 1571 respondents attend in both waves. Summary statistics for the student-level variables including information on students' characteristics and family background are displayed in Table 4. There are 77 schools in the sample. Table 5 shows the relevant summary statistics of school characteristics. This set of variables helps us to distinguish one school from another and defines the school quality. In order to measure the school quality that is crucial to this empirical study, the following aspects are taken into consideration.

1. How much market-valued skills does the school imparts to its students?

One direct way to measure the skills students acquire at their school is to assume full effectiveness of teaching at each school and then to look at the student-level curriculum content. A comprehensive curriculum includes courses on math, science, English, social science, etc. It is assumed that through its curriculum the school imparts skills that make students more productive and better rewarded in the labor market. The reasons for using math curriculum in this study is based on the following two arguments: (1) Math courses play a critical gate-keeping role in adolescents' academic careers in high school. (2) In the labor market, math skills are economically rewarded. Additionally, more-advanced math courses have greater earning effects than less-advanced ones. Rose and Betts (2004) find empirical evidence that specific high school math courses (vocational math, pre-algebra, algebra/geometry, intermediate algebra, advanced algebra, and calculus) have positive effects on earnings nearly ten years after graduation. A

varied math curriculum can explain the earnings gap between students of different ethnicities, socioeconomic statuses, and genders.

The summary statistics of the measure of student-level curriculum I use are shown in Table 4. The student-level curriculum indicator is obtained in a straightforward way from the Adolescent Health and Academic Achievement study dataset. This indicator has been constructed using a standardization procedure which enables comparisons across schools and students. It reflects both the complexity and performance expectations of curriculum content. In the appendix, there are detailed illustrations of how the indicator is constructed.

2. How do schools shape educational identity that encourages students to engage in schooling? Educational identity is an abstract concept and the intensity of educational identity q_j at each school is not directly observable. I hereby use two proxies which can capture how well students identify themselves at school. One is the school-level happiness feeling and the other, as has already been proposed by Akerlof and Kranton (2002), the school-level participation in school activities. In the Add Health dataset, all the students in each school are asked the following question towards the school they are studying in. "How strongly do you agree or disagree: You are happy to be at your school?" Additionally, all the students are asked to point out the clubs, organizations and teams in which they participate or will participate at school. I average these student-level information at each school and use them as school-level educational identity indicators.

3. Social compositions

Recent empirical studies on schools have shown that schools' social compositions (such as parents' socioeconomic status and skin color) matter to students' learning and these effects can be substantially larger than the effects of class size and teacher supply (Hanushek et al. (2006)). However, the mechanism through which the school composition enters the educational production process has not been carefully investigated. One channel proposed in this theoretical behavioral model is that the social composition at school does not directly influence learning but determines school educational identity q_j. Similarly, Akerlof and Kranton (2002) cite examples

from the sociology and education literature and suggest that schools' black and white composition influence students' attitude towards Math and English.

$$q_j = Z(I_j, F_{1j}, \ldots, F_{kj}), \text{ where } \{1, \ldots, k\} \in \text{school } j \tag{3}$$

In equation (3), I assume a simple specification of how the school-specific educational identity q_j is formed. The educational identity shaped at school *j* is not only influenced by the school organization itself I_j (i.e. teachers, administration staffs and all kinds of school arrangement) but also by the interactions among its students. Each student *i* is endowed with family impact F_{ij} and then transmits it to other students through school interactions. In empirical studies, it is still unclear which aspects of the family socioeconomic background F_{ij} significantly affect school's educational identity. Therefore, I include socioeconomic information on schools' white student proportion, migration proportion and mother college proportion in the empirical analysis to test the "educational identity" hypothesis. Tables 5 shows the summary statistics of these three social composition measures at the school-level.

4. Other factors.

Class size and teacher quality are assumed to be important determinants in the schooling process. Schools use resources to reduce the class size and attract good teachers. Smaller class size and better teacher quality concerning teaching experience, education degree and gender are expected to have a positive effect on students. This study considers that these efforts in using school resources do not only improve the teaching quality r_j at school, but also create a school community I_j, as has been shown in equation (3), which aims at establishing an ideal educational identity q_j and helps students to better fit into education. The Add Health dataset does not provide information on class size and teacher supply that can be identified at the student-level, therefore I only include relevant variables at the school-level. The school-level teacher quality can be measured as the proportions of women teachers, new teachers, experienced teachers and high degree teachers. Table 5 gives summary statistics of a set of school-level variables.

2. Empirical Model

Using the dataset described above, I attempt to answer two empirical questions. The first empirical question is how the effort e_{ij} for student i at school j is determined. Following the theoretical model, the empirical model can be specified as,

$$e_{ij} = \alpha_0 + \alpha_1 F_i + \alpha_2 A_j + \alpha_3 S_{ij}^0 + \alpha_4 IQ_i + \alpha_5 r_j + \alpha_6 q_j + \epsilon_{ij} \qquad (4)$$

The dependent variable in equation (4) is a binary variable indicating whether student i skips class in academic year 1994-95 (or year 1995-96). In the survey, students were asked "during the academic year 94-95 (or year 1995-96), how many times were you absent from school for a full day without an excuse? ". The information on whether students skip class without excuse in the academic year can be regarded as an important signal indicating the student's effort level at school. Define the student exerting high effort level $e_{ij} = e_H$ if he or she does not skip class and the student exerting low effort level $e_{ij} = e_L$ if he or she skips class. On the right hand side of the equation, the independent variables X_{ij} include: (1) the students' demographic and family background information, F_i , such as gender, race, grade, mother's education[6] and family composition; (2) the school information, A_j, such as average class and teacher supply, (3) the previous math GPA scores recorded in the school transcripts, S_{ij}^0 , (4) the IQ test, as an indicator of one's cognitive ability and scholastic aptitude, (5) the student-level math curriculum, r_j, (6) the intensity of the educational identity q_j at school j, which includes school-level of happiness and participation at clubs, teams or organizations, (7) other factors, the error term ϵ_{ij}. Assume that the error term ϵ_{ij} is independent and normally distributed, a Probit model then can be used. The hypothesized relationship is that the previous achievement S_{ij}^0, the amount of market-valued skills r_j the school imparts and the intensity of the educational identity q_j all have a positive effect on the probability of exerting high effort level in schooling. It can be expressed as follows:

$$\frac{\partial P(e_{ij} = e_H)}{\partial q_j} > 0, \qquad \frac{\partial P(e_{ij} = e_H)}{\partial S_{ij}^0} > 0, \qquad \frac{\partial P(e_{ij} = e_H)}{\partial r_j} > 0 \qquad (5)$$

---

[6] Father's education is not included in the regressors as there are more missing information in father's education than that in mother's education. Moreover, mother's education is highly correlated with father's education.

The second question is answered using a reduced form model of how the math achievement is determined. Insert equation (4) into the educational production function (2). The empirical model writes as:

$$S_{ij} = \beta_0 + \beta_1 F_i + \beta_2 A_j + \beta_3 S_{ij}^0 + \beta_4 IQ_i + \beta_5 r_j + \beta_6 q_j + \varepsilon_{ij} \qquad (6)$$

The dependent variable is the math GPA scores student i achieves in school j during academic year 1994-95 (or year 1995-96). The same set of independent variables X_{ij} as in equation (4) are applied in the estimation (6). Assume $E(X_{ij}\varepsilon_{ij}) = 0$. The hypothesized relationships are:

$$\frac{\partial S_{ij}}{\partial q_j} > 0, \qquad \frac{\partial S_{ij}}{\partial S_{ij}^0} > 0, \qquad \frac{\partial S_{ij}}{\partial r_j} > 0 \qquad (7)$$

The previous achievement S_{ij}^0, the amount of market-valued skills r_j school imparts and the intensity of the educational identity q_j all have a positive effect on the academic achievements.

Figure 2. The Educational Production Process: Direct and Indirect effects
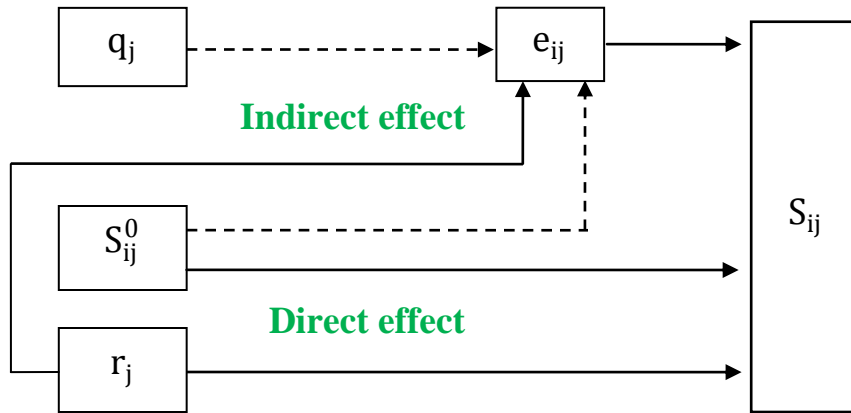
It is necessary to point out that both S_{ij}^0 and r_j enhance student's final achievement S_{ij} through two mechanisms. On the one hand, there is a direct and linear effect: S_{ij}^0 and r_j explicitly enter the education production function as is shown in (2). The current performance S_{ij} reflects both the stock and flow of the accumulation of cognitive skills. On the other hand, there is an indirect and non-linear effect: by influencing student's effort level e_{ij} as shown in (4) and (5), S_{ij}^0 and r_j

again but indirectly enter the education production function. However, the intensity of educational identity q_j does not work in the direct mechanism and only influences the student's motivation to make high effort. Figure 2 depicts the direct and indirect mechanisms in the educational production process at school. It also indicates that the behavioral approach adds more predictions to the rational expectation human capital investment model: $\frac{\partial P(e_{ij} = e_H)}{\partial q_j} > 0$, $\frac{\partial P(e_{ij} = e_H)}{\partial S_{ij}^0} > 0$, $\frac{\partial S_{ij}}{\partial q_j} > 0$.

Two potential problems can arise in the process of estimation when a pooled dataset is used. The first one is the cluster sampling problem. In this particular case, the error term is not independently anymore and identically distributed. As shown in Table 3, 1571 students are interviewed twice and in the pooled sample many students are repeatedly drawn from the same 77 participant schools. Therefore, biases will arise in the estimates of standard errors. In order to mitigate the problem, I assume that errors within groups (i.e. school-level and student-level) are correlated and then I calculate robust standard errors in the regressions. The second problem is to justify the use of the pooled dataset before combining two-wave datasets. According to empirical model (6), I do a Chow test which shows that $F (19, 5899) = 1.37$ and P-value=0.1294. At 10% significance level we cannot reject the hypothesis that "if we run two linear regressions using Wave I and II datasets separately, the coefficients from the two datasets are the same". Therefore, we can pool them together.

Part IV Results

I start with the discussion on how school resources (class size and teacher supply) affect the students' math GPA scores. Using this as a benchmark, I add information on previous math GPA scores, curriculum, school social compositions and school-level educational identity indicators to the regression of students' math GPA scores as has been discussed in equation (6). Later on, I also run the regression of students' absenteeism behavior in equation (4). This procedure helps us to disclose the direct and indirect effects in the educational process.

1. What matters: school resources or curriculum-based school effectiveness?

Column (1) of Table 6 displays the regression coefficients measuring the impact of school resources on students' math GPA scores according to regression (6). Smaller class size and lower proportion of new teachers at school are significantly associated with students' math GPA scores, while the percentage of female teachers, teachers with at least 5 years experience or with a master degree has no significant relationship with students' math GPA scores. Column (1) also tells us that students' demographic characteristics (gender, race not including Hispanic, IQ test) and family backgrounds (unstable family and mother education) influence the math GPA scores significantly.

Column (2) of Table 6, I include previous math GPA scores and student-level curriculum in the regressors. The R-square test shows that the inclusion of these two variables indeed raises the goodness of fit from 0.1306 to 0.3645 significantly. Firstly, the previous math GPA scores is crucially important. There is a strong linear relationship between the two consecutive math GPA scores S_{ij}^0 and S_{ij}. Other things being equal, an extra point previous math GPA scores (ranging from 0 to 4) will add about 0.539 to the current year's math GPA scores. Secondly, improving the quality of the school curriculum brings a considerable positive impact on the students. An extra point of the curriculum (the curriculum ranges from 0.643 to 2.589 according to our sample) will add 0.255 to the current math GPA scores.

Comparing Column (2) to Column (1) of Table 6, I find that the effects of a reduction in the average class size (from 0.013 to 0.012) and in the percentage of new teachers (from 0.515 to 0.16) remain significantly positive. After taking the previous math GPA scores and the school curriculum effects into account, the marginal effects of the students' demographic characteristics and family backgrounds drop sharply. For example, the effects of being Asian and Black become smaller and insignificant. The significant effect of being female decreases from 0.279 to 0.129. The effect of the IQ test, though small, drops from 0.019 to 0.009. The positive effects of mother's education reduce from 0.193 to 0.079. Students in an unstable family perform relatively poorer than their counterparts in a complete family and this negative impact drops from 0.249 to 0.135.

15

2. School-level educational identity: a missing dimension of school quality

Compared to the school-resources regression shown in Column (1) of Table 6, the explanatory variables in Table 7 include comprehensive information about schools. School quality is defined into two dimensions: the curriculum-based teaching quality (separately discussed in Column (2) of Table 6) and the school educational identity. In Column (1) of Table 7, I add the social composition factors (white, mother college and migration) into the regression. The effects of students' demographic characteristics and family backgrounds change accordingly. The major finding is that the proportion of students whose mother has a college degree has significant effects on the performance of the students. Other things being equal, if a student switches from a school where the mother with college degree proportion is 0 to a school where it is 1, the math GPA scores will rise by 0.393. The proportion of migrant students is also positively related to the math GPA scores but the impact is not significant.

One finding different from the existing school literature is that the proportion of white students does not significantly influence the students in school math achievement. Austen-Smith and Fryer (2005) propose a two audience signaling model where for black students signals of good academic achievements that induce high wages can be signals that induce black peer group rejection. Furthermore, Ogbu (1997) argues that there exists an "oppositional culture" on the part of black students as black students feel conflicts and tensions between their own ideals and the dominating white ideals. Akerlof and Kranton (2002) suggest that white and black students composition at school influences black students' academic attitudes and performances. To test the "oppositional culture" hypothesis among black students in American schools, I include two interaction terms "black*white student proportion" and "black*mother college proportion" in the regression. Both white student proportion and mother college proportion are good indicators of the strength of white culture. If the black is influenced by his black and white peers as the theories have suggested, we could see black students are sensitive to the social compositions in their school. However, Column (1) of Table 7 shows that black students do not perform poorer when the proportion of white students or the proportion of mother college increases. These empirical results cannot support the "oppositional culture" hypothesis in the black and white achievement gap literature.

In Column (2) of Table 7, I add the school-level happiness and school participation into the previous regression. Both of the identity indicators are significantly associated with the math GPA scores. However, the effects of mother college proportion (coefficient from 0.393 to -0.077) become insignificant. Other social composition variables such as white student proportion, migration proportion and the two interaction terms remain unimportant. The F-test (Prob > F = 0.4085) shows us that we cannot reject that the effects of the social compositions are zero. The insignificant effects of school social compositions could be due to the fact that the social composition factors are correlated with the educational identity indicators school-level happiness and participation.

Referring to specification (4), the potential correlations between the educational identity and social compositions are already defined in the school identity specification. To test the correlations, I run two regressions of educational identity indicators on the school social compositions and other school variables. The results are shown in Table 8. In Column (1) of table 8, school-level happiness is significantly related to the mother college proportion and school size. School-level happiness increases by 0.79 if the mother college proportion increases from 0 to 1. The happiness level also improves from a medium size school to a large size one by 0.143. In Column (2) of table 8, school participation is found to be significantly related to the proportion of new teachers, white student proportion, mother college proportion, migration proportion, school size and proportion in parents-teacher organization.

In Table 9, it is found that private schools have significantly higher level of happiness and school participation than public schools. Akerlof and Kranton (2002) have pointed out that private schools are able to spend considerable resources to establish school community and ensure that students identify with the school and its ideals. I compare the two types of schools in the Add health sample and the results are shown in Table 9. Private schools indeed have smaller class size and school size, higher mother college proportion and parents-teacher organization than public schools have. These differences in school arrangement can result in different emphasis on school identity.

In Column (3) of Table 7, I show the final results of a regression of math GPA scores based on the behavioral model. Both the school-level happiness and school participation are significantly important to the math GPA scores. One point improvement in the school-level happiness (the range is from 1.793 to 3.208) adds 0.241 to the Math GPA scores. One point improvement in the school-level participation in clubs, teams and organizations (the range is from 1.157 to 3.657) adds 0.152 to the Math GPA scores. Moreover, student-level math curriculum and previous math GPA scores are also significantly associated with the math GPA scores. The marginal effects of the student-level math curriculum and previous math GPA scores are 0.243 and 0.529. The estimated effects of school resources also confirm the results shown in the previous regressions. The proportion of new teachers is negatively related to the math GPA scores, the magnitude of the negative marginal effect has been reduced to 0.187. The average class size and percentage of female teachers, experienced teachers and teachers with at least a master degree are not significant determinants of students' math GPA scores.

Moreover, in Column (3) of Table 7, the marginal effects of being female, Black, an unstable family and mother with college degree are respectively 0.126, -0.096, 0.087, -0.123, 0.07. One extra point IQ test is significantly associated with an increase in math GPA scores by 0.008. I don't find any significant impact of being Hispanic and Asian. The estimated marginal effects of the students' demographic characteristics and family backgrounds drop sharply compared to our benchmark results in Column (1) of table 6. This confirms the view that there are no considerable differences in Math GPA scores among students with different races, genders and family backgrounds if the previous math GPA scores and school effects are properly controlled.

3. Does the behavioral model explain the effort level?

Table 10 shows the Probit estimation results on how students' demographic and family characteristics and school effects determine the absenteeism behavior during the academic year. Gender and IQ test scores are not significantly associated with the absenteeism behavior. Other independent variables being equal, Hispanic and Asian students are more likely to skip class by 4.4% and 8.7%. To our surprise, being Black reduces the likelihood of skipping class by 7.4% if other factors are controlled. Being in an unstable family increases the probability of skipping class by 8.1%. Mother with college degree reduces the probability of skipping class by 3.2%.

The most important results confirm the empirical hypotheses of the behavioral model in (4). Educational identity significantly influences students' absenteeism behaviors. Mother college proportion (not white student proportion and migration proportion) is found to be negatively related to the probability of skipping class, but after adding the school identity indicators the effect of mother college proportion disappears. At the mean level, the marginal effects of school-level happiness and school participation approximately reduce the chance of skipping class by 16.2% and 5.8%. Moreover, the marginal effects of improving the previous math GPA scores and student curriculum are 6.9% and 12.9% respectively. As both student-level curriculum and previous math GPA scores influence the students' effort level, this suggests that these two factors not only directly play a role in the education production function as equation (2) has described, but also indirectly influence math GPA scores through an effort choice mechanism shown in the empirical model (4) and (5). Note that the significant impact of previous math GPA scores on the effort level also informs us about the "endowment effect". The behavioral model, therefore in this perspective, explains better than a standard model.

Finally, school resources related variables also influence the absenteeism behavior. Larger class size, higher proportions of new teachers and teachers with master degree are all reported to significantly increase the chance that students skip class during the academic year.

Part V Conclusion

Schools are important social institutions for individual development. This paper provides a behavioral framework to explain the relationship between the previous educational achievement, the school quality and one's educational behavior and performance. In particular, the school quality is defined with respect to the amount of market-valued skills the school imparts and how well it establishes an educational identity. In the theoretical model, students adapt their preference of schooling according to specific educational scenarios they face. The conditions under which students fail to exert a high effort level in schooling are identified. When the previous educational achievement and school quality are poor, there exists a "low motivation and

low effort equilibrium". The school quality is defined into two dimensions: the amount of market-valued skills schools impart and how well schools cultivate an educational identity.

Using data from Add Health, I test the major hypotheses from the theoretical model and find that (1) the previous math GPA scores and the student-level math curriculum are significant determinants in math GPA scores and absenteeism behavior; (2) the educational identity indicators (school-level happiness and school participation) play an important role in shaping students' absenteeism behavior and their math GPA scores. The marginal effects of school-level happiness and school participation are 0.241 and 0.152; (3) School social composition (especially mother college proportion), school size and parents' involvement at school organization are found significantly related to the school identity indicators; (4) By contrast, the widely used variables such as average class size and teacher supply have quite limited marginal effects on education after controlling previous math GPA scores and school quality indicators.

These empirical results confirms that a behavioral empirical framework adds to the school resources approach. This proposed framework, especially the identity dimension of school quality, provides some deeper insights of how socioeconomic backgrounds widen the academic gaps between the advantaged and disadvantaged youth. This paper calls for more empirical support. Both the theoretical and empirical progress will enhance the understanding of the "poverty trap and low education" black box and help to evaluate the effectiveness of direct interventions to raise academic attainment of the disadvantaged youth.

Appendix

1. Mathematical proofs of Figure 1

For a given S_{ij}^0, the curve between "the high effort equilibrium" and "the low effort equilibrium" S_1, S_2 and S_3 can be expressed as,

$$r_j = \frac{c(IQ_i, e_H) - c(IQ_i, e_L) - q_j S_{ij}^0 \big(v(e_H) - v(e_L) \big)}{IQ_i(e_H - e_L) + q_j(v(e_H)e_H - v(e_L)e_L)}$$

When $q_j = 0$, the effect of S_{ij}^0 and q_j disappear. The student chooses $e_{ij} = e_H$ as long as

$$r_j \geq r^* = \frac{c(IQ_i, e_H) - c(IQ_i, e_L)}{IQ_i(e_H - e_L)}$$

When $r_j = 0$,

$$q_j^k = \frac{c(IQ_i, e_H) - c(IQ_i, e_L)}{\big(v(e_H) - v(e_L) \big) S_{ij}^0}, \quad k = 1, 2, 3 \text{ and } S_{ij}^0 = S_1, S_2, S_3$$

Given that $c(IQ_i, e_H) > c(IQ_i, e_L)$, $e_H > e_L$, $v(e_H) > v(e_L)$ and $v(e_H)e_H > v(e_L)e_L$,

$$\frac{\partial r_j}{\partial q_j} = -\frac{[c(IQ_i, e_H) - c(IQ_i, e_L)](v(e_H)e_H - v(e_L)e_L) + S_{ij}^0 IQ_i(e_H - e_L)\big(v(e_H) - v(e_L) \big)}{\big[IQ_i(e_H - e_L) + q_j(v(e_H)e_H - v(e_L)e_L) \big]^2} < 0$$

$$\frac{\partial^2 r_j}{\partial q_j^2} > 0$$

Therefore, at curve $S_1(S_2 \text{ or } S_3)$, r_j decreases in q_j and as q_j increases, the slope $\frac{\partial r_j}{\partial q_j}$ decreases.

2. The building of student-level curriculum indicator

According to the Academic Achievement study (http://www.prc.utexas.edu/ahaa/), student-level curriculum indicator is generated in two steps. Firstly, the textbooks used in each math course are coded. The textbook coding is based on curriculum framework developed for "*the third international math and science study*". The curriculum content is categorized into numbers, measurement, geometry, proportionality, function, relations, equations, data representation, probability, statistics, elementary analysis, validation and structure and others. This

standardization procedure enables the construction of curriculum measures that allow comparisons of students' potential curriculum exposure across different types of courses within the same school, across the same course in different schools, and accumulated across years in high school. Secondly, a summary indicator is calculated considering two aspects of a curriculum. (1) The complexity of a topic covered in a textbook. The indicator uses the International Grade Placement (IGP) rating as a weight in calculating the curriculum content in a textbook. The more complex topics contribute more to the estimate of curriculum content, the higher IGP it has. For example, curriculum "uncertainty and probability" have a higher IGP rating than curriculum "numbers" and therefore the former is weighted more than the latter to contribute to the indicator. (2) The performance expectations (PE). The indicator also relates to the review problems in the textbook, and range from routine statements of fact or execution of basic procedures to more challenging tasks such as solving a problem or communicating results. The performance expectation (PE) ratings then give more weight to material involving critical thinking in the estimation of the content of a textbook.

Table 1. Respondents' Grade Level in Each Academic Year

Add Health Survey				In school & Wave I	Wave II		
Academic Year	1991-92	1992-93	1993-94	1994-95	1995-96	1995-96	1995-96
				Grade 9	Grade 10	Grade 11	Grade 12
			Grade 9	Grade 10	Grade 11	Grade 12	
		Grade 9	Grade 10	Grade 11	Grade 12		
	Grade 9	Grade 10	Grade 11	Grade 12			

Table 2. Definitions of Variables

Variable	Description	Dataset
Student characteristics:		
Female	Female is denoted as 1, male 0.	Wave I in-home interview
Hispanic	If the race is Hispanic, 1 denotes yes, otherwise 0	
Black	If the race is Black or African America, 1 denotes yes, otherwise 0	
Asian	If the race is Asian or Pacific Islander, 1 denotes yes, otherwise 0	
White	If the race is non-Hispanic White, 1 denotes yes, otherwise 0	
IQ test scores	Add health picture vocabulary test in Wave I in-home interviews. At the beginning of the Wave I in-home interview, respondents were given the Add Health Picture Vocabulary Test, a computerized and abridged version of the Peabody Picture Vocabulary Test. In this test, the interviewer reads a word and the respondent selects the illustration that best fits its meaning. There are 87 items on the AHPVT, and raw scores have been standardized by age. It provides an estimate of verbal ability or scholastic aptitude.	
Grade10-12	Grade10 (or Grade11) mean the student is in the 2nd (or 3rd) year at school. Grade 12 include those who have been at school for 4-6 years.	the Adolescent Health and Academic Achievement study
Family background:		
Unstable family	If the subject does not live with either biological father or mother or both, we define it as 1, otherwise 0.	Wave I in-home interview
Mother with a college degree	If mother with at least a college degree, define it as 1, otherwise 0.	
Educational information:		
Math GPA	The math scores in the survey academic year 1994-95 (Wave I) or 1995-96 (Wave II). It is the average GPA of the math courses taken in that particular year. All the test results are recorded according to each student's official transcript collected from his or her school.	the Adolescent Health and Academic Achievement study
Previous math GPA	It is the previous math test scores on the students' official transcript prior to math GPA in the survey year.	
Student-level curriculum	An indicator of school quality, which measure how much math skills and knowledge students are exposed to in a particular year's math course-taken. The range of the indicator is from 0.6432114 to 2.588921. The higher, the more valued.	
Absenteeism behavior	An indicator of how much effort one put into study. Students were asked "during the academic year 94-95 (or year 95-96), how many times were you absent from school for a full day without an excuse? " If the answer is at least one time, denote it 1; otherwise, 0.	Wave I and II in-home interview
School-level information:		
Private	School type: if private assigns as 1, public 0	
School size	1-3 which indicate small(1-400), medium(401-1000), large(1001-4000)	
Average class	The average class of school. It ranges from 12 to 38.	
New teacher	The proportion of teachers who are new in the school.	
Women teacher	The proportion of teachers who are women in the school.	
Experienced teacher	The proportion of teachers who have at least 5 years experience in the school.	
Master or higher	The proportion of teachers who have master or higher degree in the	

teacher	school.	school administration questionnaire
White student proportion	The proportion of students who are white	
Dress code	Whether students must obey a dress code at 9th grade. Yes, 1; No, 0	
Proportion parents-teacher organization	% of children with family in parent organization	
Mother college proportion	The proportion of students whose mother has at least a university degree. Using information from all the students at school, the average level at each school is calculated.	In-School Questionnaire
Migrant school proportion	The proportion of students whose parents are both not born in the US. Using information from all the students at school, the average level at each school is calculated.	
School-level student feel happy at school	It indicates the intensity of the educational identity or the salient of an educational identity at school level. The question is "How strongly do you agree or disagree: You are happy to be at your school? " The answer is on a 0-4 scale from strongly disagree to strongly agree. Using this information from all the students at school, the average level at each school is calculated.	
School-level student participation at school clubs, teams and organizations.	It indicates the intensity of the educational identity or the salient of an educational identity at school level. The students are asked to give the list of clubs, organizations and clubs they participate or will participate. Using this information from all the students at school, the average participation level at each school is calculated.	

Table 3. Description of Sample Structure

	Number of Observation
Only in Wave I	1495
Only in Wave II	1300
Both in Wave I & II	1571*2=3142
Total Sample	5937
Number of participation schools	77

Table 4. Summary Statistics of Student Characteristic, Family and Academic Information

Variable	Wave I (N=3066)	Wave II (N=2871)	Pooled (N=5937)	Min	Max
Individual Characteristics:					
Female	0.53	0.546	0.538	0	1
Hispanic	0.145	0.126	0.136	0	1
Black	0.186	0.197	0.192	0	1
Asian	0.12	0.105	0.113	0	1
Grade 10	0.378	0.344	0.362	0	1
Grade 11	0.399	0.38	0.39	0	1
Grade 12	0.223	0.276	0.248	0	1
IQ test	102.309 (13.528)	102.777 (14.478)	102.536 (13.996)	10	133
Family Background:					
Unstable Family	0.364	0.347	0.356	0	1
Mother with college degree	0.307	0.333	0.32	0	1
Educational information:					
Math GPA scores	2.169 (1.16)	2.171 (1.178)	2.17 (1.169)	0	4
Skip class without excuse	0.352 (0.478)	0.393 (0.488)	0.372 (0.483)	0	1
Previous math GPA scores	2.4 (1.104)	2.359 (1.133)	2.38 (1.118)	0	4
Curriculum	1.024 (0.192)	1.018 (0.192)	1.021 (0.192)	0.643	2.589

Table 5. Summary Statistics of School Characteristics (Number of Schools=77)

Variable	Mean	Std. Dev.	Min	Max
Private School	0.104	0.307	0	1
Average Class Size	25.169	5.401	12	38
New Teacher	0.101	0.137	0	0.99
Women Teacher	0.584	0.161	0.23	0.96
Experienced Teacher	0.642	0.239	0	0.98
Teacher with Master or Higher Degree	0.473	0.239	0	0.95
White Student Proportion	0.651	0.321	0	1
Mother College Proportion	0.301	0.138	0.061225	0.829586
Migrant Student Proportion	0.087	0.149	0	0.81658
Feel Happy	2.531	0.249	1.793443	3.208333
Participation at School Clubs/Teams/Organizations	2.276	0.513	1.156702	3.657408
School Size	2.234	0.705	1	3
Dress Code	0.701	0.461	0	1
Proportion parents-teacher organization	0.214	0.239	0	1

Table 6. School Quality and Math Performance (I)

	Math GPA scores			
	Coefficient			
	(1)		(2)	
Female	0.279***	(0.027)	0.129***	(0.023)
Hispanic	-0.121	(0.147)	-0.062	(0.063)
Black	-0.245***	(0.094)	-0.083	(0.053)
Asian	0.301**	(0.126)	0.104	(0.074)
IQ test	0.019***	(0.001)	0.009***	(0.001)
Unstable family	-0.249***	(0.046)	-0.135***	(0.031)
Mother with college degree	0.193***	(0.054)	0.079**	(0.036)
Average class size	-0.013*	(0.006)	-0.012**	(0.005)
new teachers	-0.515***	(0.136)	-0.16*	(0.092)
Women teachers	0.099	(0.233)	0.043	(0.166)
teachers at school 5 years or more	-0.056	(0.156)	-0.041	(0.116)
teachers with master or higher degree	0.05	(0.18)	-0.013	(0.117)
Previous math GPA scores	--		0.539***	(0.023)
Curriculum	--		0.255**	(0.1)
Estimation	**OLS**		**OLS**	
Number of clusters (individuals)	**4366**		**4366**	
Number of clusters(schools)	**77**		**77**	
Prob > F	**0.0000**		**0.0000**	
R-square	**0.1306**		**0.3645**	
Number of observations	**5937**		**5937**	
R-Square test	**(1) Previous math GPA scores = 0** **(2) Curriculum = 0** **F(2, 4365) = 300.37** **Prob > F = 0.0000** **Conclusion: Statistically Significant Increase in R-Square**			

Additional to the coefficients of the estimations, in the table, robust standard error are shown in the parentheses.
*** indicates significant level at 1%, ** at 5%, * at 10%. Grade and wave information is also included in the regressions.

Table 7. School Quality and Math Performances (II): Adding Identity Indicators

Variable	Math GPA scores		
	Coefficient (Robust Std. Err.)		
	(1)	(2)	(3)
Female	0.129*** (0.022)	0.127***(0.022)	0.126***(0.022)
Hispanic	-0.024 (0.052)	-0.037 (0.053)	-0.052 (0.062)
Black	-0.142 (0.137)	-0.174 (0.126)	-0.096**(0.044)
Asian	0.105 (0.064)	0.1 (0.063)	0.099 (0.061)
IQ test	0.008*** (0.001)	0.008***(0.001)	0.008***(0.001)
Unstable family	-0.128***(0.031)	-0.122***(0.031)	-0.123***(0.031)
Mother with college degree	0.07* (0.036)	0.076** (0.035)	0.07** (0.035)
Average class size	-0.008 (0.005)	-0.001 (0.006)	-0.004 (0.006)
new teachers	-0.286** (0.12)	-0.19 (0.117)	-0.187* (0.102)
female teachers	0.085 (0.188)	0.065 (0.168)	0.146 (0.164)
teachers at school 5+ years	-0.041 (0.11)	0.026 (0.097)	0.031 (0.106)
teachers with master+ degree	-0.062 (0.111)	-0.041 (0.102)	-0.04 (0.101)
Previous math GPA scores	0.532***(0.023)	0.528***(0.024)	0.529***(0.024)
Curriculum	0.271***(0.103)	0.264***(0.102)	0.243** (0.099)
White student proportion	0.003 (0.117)	-0.021 (0.098)	--
Mother college proportion	0.393***(0.148)	-0.077 (0.182)	--
Migration proportion	-0.2 (0.136)	-0.147 (0.15)	--
Black*white student proportion	0.097 (0.161)	0.223 (0.147)	--
Black*Mother college proportion	0.059 (0.388)	0.013 (0.34)	--
School-level feel happy	--	0.309***(0.099)	0.241***(0.091)
School participation	--	0.14*** (0.054)	0.152***(0.044)
Estimation	**OLS**	**OLS**	**OLS**
Number of clusters (individuals)	**4366**	**4366**	**4366**
Number of clusters(schools)	**77**	**77**	**77**
Prob > F	**0.0000**	**0.0000**	**0.0000**
R-square	**0.3660**	**0.3699**	**0.3692**
Number of observations	**5937**	**5937**	**5937**

Additional to the coefficients of the estimations, in the table, robust standard error are shown in the parentheses.
*** indicates significant level at 1%, ** at 5%, * at 10%.

-- F test: (1) White student proportion = 0; (2) Mother college proportion = 0; (3) Migration proportion = 0; (4) Black*white student proportion = 0; (5) Black*Mother college proportion = 0. Results: $F_{(5, 4365)}$ =1.01 and Prob > F = 0.4085. We cannot reject that the coefficients are zero, therefore I exclude the social composition variable in the regression in Column (3).

Table 8. School factors on School-level Happiness and Participation

Variable	School-level Happiness		School Participation	
	(1)		(2)	
Private school	0.008	(0.097)	0.113	(0.235)
Average class size	0.001	(0.006)	-0.004	(0.01)
new teachers	-0.071	(0.145)	-0.941***	(0.296)
female teachers	0.27	(0.205)	-0.363	(0.362)
teachers at school 5 years or more	-0.001	(0.095)	0.005	(0.166)
teachers with master or higher degree	-0.088	(0.126)	-0.094	(0.154)
White student proportion	0.056	(0.091)	0.333**	(0.134)
Mother college proportion	0.79***	(0.249)	1.467***	(0.416)
Migration proportion	0.394	(0.196)	-0.69***	(0.25)
School size	-0.143***	(0.045)	-0.276***	(0.068)
School dress code	-0.011	(0.06)	0.068	(0.084)
Proportion parents-teacher organization	-0.175	(0.16)	0.798***	(0.188)
Estimation	**OLS**		**OLS**	
Prob > F	**0.0002**		**0.0000**	
R-square	**0.3099**		**0.5846**	
Number of schools	**77**		**77**	

Additional to the coefficients of the estimations, in the table, robust standard error are shown in the parentheses.
*** indicates significant level at 1%, ** at 5%, * at 10%.

Table 9. Private and Public Schools

	Private School	Public School	Mean comparison
School-level Happiness	2.766	2.504	+***
School-level Participation	2.837	2.211	+***
Average class size	20.625	25.696	- ***
New teachers	0.196	0.09	+**
Female teachers	0.59	0.583	+
Teachers at school 5 years or more	0.464	0.663	-**
Teachers with master or higher degree	0.39	0.483	-
White student proportion	0.656	0.651	+
Mother college proportion	0.468	0.282	+***
Migration proportion	0.119	0.083	+
School size	1.5	2.319	-***
School dress code	0.875	0.681	+
Proportion parents-teacher organization	0.324	0.202	+*
Number of Observations	8	**69**	

+ indicates private school>public school; - indicates public school>private school
*** indicates significant level at 1%, ** at 5%, * at 10%.

Table 10. School Quality and Absenteeism

Variable	Skip class (yes= 1, no= 0)			
	Coefficient	Robust Std. Err.	Mean	dF/dx
Female#	-0.011	0.03	--	-0.004
Hispanic#	0.117*	0.064	--	0.044
Black#	-0.203***	0.061	--	-0.074
Asian#	0.227**	0.108	--	0.087
IQ test	0.001	0.001	102.535	0
Unstable family#	0.216***	0.042	--	0.081
Mother with college degree#	-0.085**	0.042	--	-0.032
Average class size	0.026***	0.006	27.02	0.01
new teachers	0.275**	0.111	0.105	0.103
female teachers	0.25	0.182	0.541	0.094
teachers at school 5 years or more	-0.103	0.24	0.63	-0.039
teachers with master or higher degree	0.402**	0.161	0.422	0.15
Previous math GPA scores	-0.184***	0.028	2.38	-0.069
Curriculum	-0.346***	0.126	1.021	-0.129
School-level feel happy	-0.434***	0.125	2.531	-0.162
School Participation	-0.154**	0.071	2.046	-0.058
Estimation	Probit			
Number of clusters (individuals)	4366			
Number of clusters (schools)	77			
Prob > chi2	0.0000			
Pseudo R-square	0.0829			
Number of observations	5937			

*** indicates significant level at 1%, ** at 5%, * at 10%.

(#) dF/dx is for discrete change of dummy variable from 0 to 1; other dF/dx calculates the marginal effects at the means of the independent variables.

References

1. Akerlof, George and Rachel Kranton (2002), "Identity and Schooling: Some Lessons for the Economics of Education", Journal of Economic Literature, Vol. XL, p.1167-1201

2. Atkinson, Tony (2003), "Income Inequality in OECD Countries: Data and Explanations CESifo Economic Studies", Vol. 49, p.479-513

3. Austen-Smith, David and Roland Fryer (2005), "An Economic Analysis of `Acting White", Quarterly Journal of Economics, May, p.551-583

4. Bandura, A. (1986), Social foundations of thought and action: A social cognitive theory, Englewood Cliffs

5. Bandura A. (2001) "Social cognitive theory: an agentic perspective," Annual Review of Psychology, 52, 1-26

6. Benabou, Roland (1993) Workings of a City: Location, Education, and Production, Quarterly Journal of Economics, Vol. 108, No. 3, p. 619-652

7. Burtless, Gary, (1996), "Does Money Matter?" Washington, DC: Brookings Institution

8. Dominitz, Jeff and Charles Manski, (1996), "Eliciting Student Expectations of the Return to Schooling", The Journal of Human Resources, p.1-26

9. Gary S. Becker (1964), "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education", Chicago, University of Chicago Press

10. Hanushek, Eric A. (2003), "The Failure of Input-based Schooling Policies", Economic Journal 113, no.485 (February): F64-F98

11. Hanushek, Eric A. and Welch Finis, Edit(2006), "Handbook of the Economics of Education", Volume 2 Chapter 14, 15 and 17, Amsterdam, North-Holland

12. Heckman, James, Jora Stixrud and Sergio Urzua (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", Working paper 12006, National Bureau of Economic Research

13. Hoxby, Caroline, (2000) "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Papers 7867, National Bureau of Economic Research

14. James, William (1890), "The Principles of Psychology", 2 vols. (1890) Dover Publications 1950, vol. 1

15. Ji, Yuemei (2008), "Identity and Educational Choice: An Behavioral Approach", University of Leuven, CES Discussion Paper

16. Rose, Heather and Julian R. Betts (2004), "The effect of High School Course on Earnings", Review of Economics and Statistics, May 2004, 86(2): p. 497-513

17. Kozol, Jonathan (1991), "Savage Inequalities", New York: Crown

18. Ogbu, John (1997), " Beyond Language: Ebonics, Proper English, and Identity in a Black-American Speech Community," mimeo, University of California Berkeley, Department of Anthropology