

MISBEHAVIOR, EDUCATION, AND LABOR MARKET OUTCOMES

Carmit Segal
University of Zurich

Abstract

Using data on young men from the National Education Longitudinal Survey, this paper investigates the relationship between childhood misbehavior and later education and labor market outcomes. The main finding is that eighth-grade misbehavior is important for earnings over and above eighth-grade test scores. Moreover, controlling for educational attainment, childhood misbehavior is associated with earnings at all educational levels, whereas achievement test scores are related to earnings only for young men with postsecondary degrees. Possible explanations for the association between eighth-grade misbehavior and economic success are explored. (JEL: J24, J31, J13, I21)

1. Introduction

In recent years, researchers have noted that socially productive skills include not only the traditionally investigated cognitive skills but also noncognitive ones. This broader concept of human capital has naturally influenced the research of labor economists and macroeconomists both. The macroeconomic literature investigates how personality traits such as work ethic, patience, and the individual's discount rate affect economic growth. Thus, Becker and Mulligan (1997) argue that the evidence on cross-country differences in growth and on within-country consumption and earning inequalities is consistent with individuals having varying discount rates (and with their model, in which individuals can decide their discount rate). Doepke and Zilibotti (2008) explain the rise of the middle class and the decline of the aristocracy in the aftermath of the

The editor in charge of this paper was Fabrizio Zilibotti.

Acknowledgments: This paper is part of my Ph.D dissertation at Stanford University. I thank my advisors Ed Lazear, Muriel Niederle, and Ed Vytalil for their encouragement, useful suggestions, and numerous conversations. I thank the editor, Fabrizio Zilibotti, and three anonymous referees for their helpful comments and suggestions and Joao de Mello, Vinicius Carrasco, Liran Einav, Christina Gathmann, Avner Greif, James Heckman, Larry Katz, Frank Levy, Susanna Loeb, Tom MaCurdy, Aprajit Mahajan, Tatiana Melguizo, Amalia Miller, John Pencavel, Luigi Pistaferri, Esteban Rossi-Hansberg, Daniel Tsiddon, Karine van der Beek, Pierre-Olivier Weill, Frank Wolak, Nese Yildiz and multiple discussants and seminar participants for helpful comments and insights on earlier drafts. Special thanks to Ilona Berkovits for access to and assistance with the data.

E-mail: carmit.segal@econ.uzh.ch

Industrial-Revolution as a result of different investment in skills. The authors suggest that while upper-class families cultivated in their offspring a taste for leisure, middle-class families instilled values of patience and hard work that enabled their offspring to succeed economically in the industrialized world. In support of the view that religious belief influences traits (e.g., a work ethic) important for economic growth, Barro and McCleary (2003) document that, across countries, growth is positively correlated with the prevalence of beliefs in heaven and hell. In labor economics, a new literature has emerged that documents the importance of noncognitive skills to various economic outcomes (see, e.g., Heckman, 2000; Bowles et al., 2001a; Cawley et al., 2001; Heckman and Rubinstein, 2001; Jacob, 2002; Carneiro and Heckman, 2003; Persico et al., 2004; Kuhn and Weinberger, 2005; Heckman et al., 2006; Lindqvist and Westman, 2011). The noncognitive traits investigated to date, in this literature, include social and leadership skills and various personality traits such as aggression, externality, self-esteem, and locus of control. Whereas such traits are important from a psychological view point, the ones that were deemed important for economic growth (i.e., self-control, patience, and hard work) were not investigated. This paper contributes to the literature on the relationship between noncognitive skills and economic success by focusing on proxies for these exact traits.¹ Moreover, whereas the previous literature used measures that rely on self-reporting – that in some cases were reflective of, or even concurrent with, the outcomes of interest² – this paper uses pre-market eighth-grade teacher evaluations to measure noncognitive skills.

The analysis in this paper is based on young men in the National Educational Longitudinal Survey (NELS) data set;³ this panel data set spans a period of twelve years, beginning when participants were in eighth grade. It contains information on participants while they were still in school as well as information on educational

1. The macroeconomic literature on quasi-hyperbolic discounting focuses on another aspect of discounting: the consequences of individuals' tendency to discount the future to a greater extent when making trade-offs that involve the present (for an excellent summary of the theory and evidence see DellaVigna, 2009). I know of no direct evidence connecting the noncognitive measures used in this paper to the tendency to be hyperbolic, though some indirect evidence may be found in Ameriks et al. (2007).

2. An exception is Lindqvist and Westman (2011), who use psychologist evaluations of the leadership skills of potential recruits to the Swedish army as a measure of noncognitive skills.

3. A number of papers have cautioned against inferences made from female earnings to offered wages, because of severe selection problems. Specifically relevant for this paper is the work of Mulligan and Rubinstein (2008), who suggest that selection is an important determinant of female wages and that relationships between (cognitive) skills and wages for women are nontrivial. Unfortunately, there are no papers that investigate, for females, the role of noncognitive skills in either labor market selection or educational attainment choices. Issues related to female labor supply and educational choices are therefore beyond the scope of this paper. That being said, girls do behave significantly better than boys on all measures other than absenteeism. A possible explanation for girls' higher absenteeism rates is biological; in particular, menstruating young women may be more likely to miss school (for documentation in adults, see Ichino and Moretti, 2009).

attainment and labor market outcomes. Standardized achievement tests were administered to all participants, whose parents, teachers, and school administrators were also surveyed. Teacher evaluations of student behavior in eighth grade are used to create a pre-market measure of participants' behavior, thereby avoiding possible problems associated with self-reporting. The teachers evaluated their students on five attributes: absenteeism, disruptiveness, inattentiveness, tardiness, and homework completion. Thus, the evaluations provide a broad measure of student behavior in the classroom.⁴

In this paper, I investigate the relationship between misbehavior in eighth grade and educational and labor market outcomes. After controlling for eighth-grade test scores and family characteristics, I find that misbehavior in eighth grade is negatively correlated with educational attainment. Moreover, after controlling for eighth-grade test scores, I find that misbehavior in eighth grade is negatively correlated with earnings of young men who are 26–27 years old. Furthermore, once individuals' education levels are controlled for, misbehavior in eighth grade is associated with lower earnings for young men of all education levels, whereas performance on the eighth-grade achievement tests is associated with earnings only for young men with postsecondary degrees.

The forms of misbehavior investigated in this paper are minor in comparison to the criminal behavior, drug abuse, violence, and aggression that have received so much attention in the literature (for an excellent summary of the research on antisocial behavior in children and youth, see Dodge, 2006). Nevertheless, classroom misbehavior – especially when it reflects the absence of traits that are important in the labor market – may have severe consequences.⁵ Indeed, substantial resources in the area of organizational behavior are devoted to developing tools that enable employers to identify individuals that will display counterproductive work behavior, such as absenteeism, tardiness, disciplinary problems, theft, sexual harassment, accidents, sabotage, drug or alcohol abuse, violence, or white collar crime (see Roberts et al. (2007) and there the work cited). There is evidence (Ones et al., 2007) that individuals

4. I discuss the reliability of teacher evaluations in Section 3.

5. It seems that the good behavior required in school is not unlike the good behavior required in the workplace. In both settings, students/employees are required to be on time and not miss school/work without a valid excuse. In both, students/employees are assigned work (homework/job tasks) that needs to be completed by a certain deadline. One's standing with the teacher/employer is determined by the quality of one's work and by the behavior one displays in the classroom/workplace. Furthermore, theories from sociology of education suggest that schools are not only where knowledge is acquired but also where young members of society are prepared to take part in the political and economic spheres of society (see Parsons, 1959; Bowles and Gintis, 1976; Dreeben, 1976). Hence, these theories imply that one task of schools is to teach students how to behave in the workplace. Individuals who have failed to acquire these skills (i.e., who have not learned to follow these rules of society) by the eighth grade may not be able to acquire them in the future and so may have worse outcomes.

who display counterproductive work behavior are less conscientious.⁶ Roberts et al. document that low conscientiousness at age 18 predicts counterproductive work behavior at age 26.⁷ These authors also document that adolescent conduct disorder (which includes antisocial behavior in addition to violation of school rules) correlates with counterproductive work behavior and with low levels of conscientiousness at age 18. Barbaranelli et al. (2003) show that low conscientiousness in junior high correlates with both externalizing behavior (which includes misbehavior in school) and low grades in school.

In economics, traits like self-control and inclination to plan are described by the discount rate. The evidence in economics suggests that eighth-grade misbehavior may relate to low discount rates. Specifically, Castillo et al. (2011) elicit discount rates of the entire eighth-grade student population of one school district in Georgia by offering the eighth-graders choices between receiving \$49 after a month and receiving a larger amount of money that varied (between \$50.83 and \$98.02) after seven months. The authors find that eighth-graders with higher discount rates have higher rates of disciplinary referrals in both the eighth and ninth grades.⁸

Taken together, the economics and psychology literatures suggest that individuals with high discount rates are more likely to be conscientious and that misbehavior is negatively correlated with this trait. Hence, the favored explanation in this paper is that eighth-grade misbehavior is indicative of inappropriate behavior displayed later on in school and at the workplace. Thus, misbehaving students are boys with low discount rates, which in turn are related to low levels of conscientiousness (i.e., low noncognitive skills).

Section 2 is devoted to investigating mechanisms that may be responsible for the associations found in the data. I present a model that follows the skill-weights approach suggested by Lazear (2009). In the model, individuals are endowed with both cognitive and noncognitive human capital. Firms use both forms of human capital to produce output, but the relative importance of each varies across firms. Before starting to work, an individual must decide whether or not to invest in schooling – a technology that enhances one’s cognitive skills – while taking into account the

6. In psychology, the term “conscientious” is used to describe the tendency to follow rules and to be task- and goal-directed, inclined to plan, and self-controlled (Roberts et al., 2004). Ones et al. (2007) summarize evidence on the strong relationship between conscientiousness and other job performance criteria.

7. This accords with findings that conscientiousness is a personality trait that, more than any other, seems to persist throughout one’s life (see, Roberts et al., 2004, for a review of the empirical evidence).

8. In addition, Castillo et al. (2011) find that “impatience patterns” by race and gender are similar to those displayed by eighth-grade misbehavior. Specifically, they find that boys and African-Americans are more likely to be impatient.

costs and expected benefits. Interpreting the eighth-grade behavior and test scores as proxies for the stock (accumulated by that grade) of noncognitive and cognitive human capital, respectively, the model is able to explain the empirical findings. Lower schooling levels are the result of lower expected returns to schooling and (possibly) of the higher costs associated with lower levels of cognitive and noncognitive skills. Misbehavior is correlated also with lower educational attainment because it implies higher discounting of the returns to schooling. Lower earnings are the result of lower productivity, which is directly associated with one's cognitive and noncognitive human capital. I then extend the model to provide a mechanism by which noncognitive skills can affect cognitive human capital accumulation. In the extended model, individuals can enhance their cognitive human capital at each level of schooling by exerting costly effort. Thus, those who value the future less (i.e., those with low noncognitive skills) invest less effort in school and hence accumulate less cognitive human capital; as a result, they earn less. The existence of individual differences in discount rates therefore generates cognitive human capital differentials in addition to those stemming from different schooling levels and different endowments.

The extended model also emphasizes that one should not expect all noncognitive traits to have the same relationship with outcomes. In particular, it suggests that misbehavior – through its connection to the discount rate – may result in lower cognitive human capital, which will lead to even lower earnings than already predicted by the lower productivity associated with misbehavior. This relation between cognitive human capital and misbehavior need not hold for other noncognitive traits. Thus, for example, we expect traits like social skills and leadership to be correlated with earnings through their effect on productivity in jobs where these traits are essential, though they may not affect earnings in other jobs. Implicitly, then, the extended model suggests that distinctions should be made between different noncognitive traits and that (for the purpose of analysis) they should not necessarily be bundled together.

Alternative interpretations are investigated in Section 5. First, the associations just summarized may not mean that classroom misbehavior itself causes worse outcomes. It is possible that misbehavior is merely a by-product of an omitted variable or of a variable measured with an error that is causally related to outcomes. Besides the discount rate, candidate variables include preference parameters (or expectations) and low cognitive skills. Specifically, it is possible that the correlations in the data are driven by boys' tastes for occupations or schooling irrespective of their (cognitive or noncognitive) skills. This mechanism is similar to that of the extended model. Though in this mechanism, tastes and expectations (and not the discount rate) change, in practice, the returns from behaving well in school. Boys who value neither schooling nor the knowledge acquired thereby (either in itself or as a means to an end) may decide not to invest effort and to misbehave. This would lead them to accumulate

less cognitive human capital and hence reduce their future earnings (and potentially their educational attainment). I find no evidence that such channels exist. Even after controlling for eighth-graders' stated expected occupation at age 30 or their stated expected education levels, the results remain intact.

A second possible explanation is that misbehavior in school is merely a by-product of low cognitive skills (i.e., those who have a hard time understanding the material taught in school misbehave). This is a valid criticism of all the literature on noncognitive skills. If success were determined solely by cognitive skills then it follows that – because any psychological measure may be affected negatively by failure in school or in the workplace – worse psychological traits would imply low cognitive skills. It seems impossible that eighth-grade test scores are perfect measures of cognitive ability and that misbehavior is simply a by-product of low cognitive ability. If that were the case then there would not have been, after controlling for eighth-grade test scores, any associations between eighth-grade misbehavior and later outcomes. Thus, one can only argue that cognitive ability is measured with error by eighth-grade test scores and misbehavior both. However, the findings reported here suggest that to believe this requires one to believe also that, for less-educated individuals, eighth-grade misbehavior is not just a better proxy (than eighth-grade test scores) for cognitive skills but in fact is the only (viable) proxy.

Even if the findings reported in this paper imply that misbehavior in eighth grade results in lower educational attainment and earnings, we still need to ask whether it is possible for an isolated incident of misbehavior in the eighth grade to be the causes of worse outcomes more than a decade later – that is, without individuals displaying inappropriate behavior (in school and possibly in the workplace) during subsequent years. If it is possible, then misbehavior in eighth grade would not necessarily indicate one's type (i.e., discount rate or personality traits). To the best of my knowledge, there is no evidence in the literature suggesting that the eighth grade is special in terms of educational progress and that misbehavior at this stage dooms one for life. A more plausible explanation is that being inattentive in class, missing school, and not doing one's homework (i.e., misbehaving) may result in less accumulated knowledge. After controlling for family background characteristics, school fixed effects, and eighth-grade test scores, I find that eighth-grade misbehavior is associated with lower tenth-grade test scores. However, neither one's background nor one's above-average knowledge in eighth grade mitigates this effect. It thus seems that those who misbehave not only accumulate less knowledge but they also continue to misbehave.⁹ So we return to the

9. The results in Segal (2008) suggest that this is indeed the case. Using the NELS teacher evaluations of student behavior in 1988 (when participants were in eighth grade) and 1990 (when most participants were in tenth grade), the author shows that boys' behavior in school is fairly stable at least after the eighth grade. That work also shows that even though family and school characteristics, as well as the incentives for good behavior provided at home and in school,

explanation that those who misbehave in eighth grade keep misbehaving later on in school and still later in the workplace, and therefore earn less.

The paper proceeds as follows. Section 2 presents the model and Section 3 describes the NELS data set. Section 4 gives a detailed analysis of the relationship between misbehavior in eighth grade and outcomes. Section 5 explores the channels that could account for the association between misbehavior and outcomes and offers some interpretations. Section 6 concludes.

2. The Model

In this section I present a model demonstrating how cognitive and noncognitive skills may affect economic outcomes and then discuss its relationship to possible empirical specifications. The purpose is to understand what relationships may exist in the data between economic outcomes and eighth-grade misbehavior and test scores. To achieve this goal, I do not write the most general model possible. Instead, I make specific assumptions that are motivated by evidence in the literature and the data available in the NELS.

2.1. The Basic Model

Individuals live for two periods. In the first period, an individual must decide whether to invest in schooling or not; in the second period, he works. Individuals are endowed with a two-dimensional skill vector $(A, \theta) \in \mathbb{R}_{++}^2$, where A and θ denote (respectively) the individual's cognitive and noncognitive skills. Higher values of A and θ are associated with better outcomes. Given the data, a sensible interpretation of the endowments θ and A is that θ represents noncognitive human capital accumulated by the eighth grade while A represents the corresponding cognitive human capital.¹⁰

When deciding whether to invest in education, each individual takes into account the associated benefits and costs. Benefits could be of two kinds: an increase in cognitive human capital (denoted by H_C) and/or an increase in noncognitive human capital (denoted by H_{NC}). When discussing the link to the data, the decision to invest in schooling is interpreted as the decision to obtain a postsecondary degree. It therefore seems natural to assume that schooling increases an individual's cognitive human capital. However, since research (Segal, 2008) indicates that boys' behavior in school is quite stable at least after the eighth grade, I assume that postsecondary

are important determinants of behavior in the classroom neither the cross-sectional variation in behavior nor the variation in behavior over time can be attributed to these covariates.

10. This interpretation assumes that the production process of future human capital depends directly on the current stock – not the initial levels – of human capital.

education does not change one's noncognitive human capital; hence $H_{NC} = \theta$.¹¹ The production function of cognitive human capital is a function of A , θ , and S ; where S denotes schooling. For simplicity I assume that $S \in \{0, 1\}$, where $S = 1$ if the individual acquired schooling at a cost $C(A, \theta)$ and $S = 0$ otherwise. The costs associated with the schooling investment are (weakly) decreasing in cognitive and noncognitive skills; that is, $C_A \leq 0$ and $C_\theta \leq 0$, respectively.¹²

Firms use both cognitive and noncognitive human capital to produce output. Following the skill-weights approach suggested by Lazear (2009), I assume that firms are heterogeneous in the sense that they use different weights of the two skills in production. Thus I assume that there is a continuum of firms each characterized by λ , a random variable with well-behaved density and cumulative distribution functions. The output at firm i of a worker with a skill set (H_C, H_{NC}) is given by

$$y_i = \lambda_i H_C + (1 - \lambda_i) H_{NC}, \quad (1)$$

with $0 \leq \lambda_i \leq 1$. The term λ_i reflects the fact that firms may weigh the two skills differently, from which it follows that individuals may have different productivities in different firms.

In the second period (i.e., the working period), each worker receives one job offer and decides whether to accept or reject it. Because the situation is one of an ex post bilateral monopoly, the wage is determined by some equilibrium bargaining game. For tractability I assume that the wage is determined by Nash bargaining, although the results remain qualitatively the same as long as some rent splitting occurs. Denote the firm at which the worker is employed during the second period as firm 1. The worker's output at firm 1 is: $y_1 = \lambda_1 H_C + (1 - \lambda_1) H_{NC}$. The worker's outside option is unemployment, which is normalized to have value 0. Hence Nash bargaining implies that, in the second period, the worker's wage is given by $w = 0.5[\lambda_1 H_C + (1 - \lambda_1) H_{NC}]$. Assuming that $A \geq 0$ and $\theta \geq 0$, all individuals decide to work in the second period.

The production function of cognitive human capital is given by

$$H_C = \begin{cases} g(A, \theta) & \text{if } S = 1 \\ A & \text{if } S = 0. \end{cases} \quad (2)$$

11. This does not preclude the possibility that schooling and parental investment at earlier ages do affect noncognitive and cognitive human capital accumulation or that postsecondary education increases other, noncognitive skills such as social and leadership skills. This assumption is in line with the interpretation that eighth-grade misbehavior proxies for noncognitive human capital (accumulated by the eighth grade) and not for raw noncognitive ability.

12. Because higher values of both A and θ contribute to higher productivity (i.e., are associated with better outcomes), the signs on the partial derivatives of the cost function are opposite to the usual signs.

To guarantee that indeed $g(A, \theta) > A$ for all A and θ , I assume that $g_A > 1$; this implies that cognitive human capital is increasing in A . In addition I assume that $g_\theta \geq 0$, which means that those with higher θ acquire (weakly) higher cognitive human capital for the same level of A . Thus, schooling results in higher cognitive human capital for all values of cognitive and noncognitive skills, yet the increase in cognitive human capital due to schooling is higher for those with higher skills. Because the worker must decide on his human capital investment strategy *before* observing the realizations of λ , his problem is

$$\max_{S=\{1,0\}} \beta Ew(S) - C; \quad (3)$$

where β is the discount rate (given that the costs are in the present and the benefits are in the future) and $Ew(S) = 0.5[E\lambda_1 H_C + (1 - E\lambda_1)\theta]$ is the expected wage in period 2 (after substituting $H_{NC} = \theta$). Since $E\lambda_1 \leq 1$, the expected wage is nonnegative for all workers.

In the first period, the individual chooses whether or not to invest in schooling. His expected wage is given by

$$Ew = \begin{cases} \frac{1}{2} [E\lambda_1 g(A, \theta) + (1 - E\lambda_1)\theta] & \text{if } S = 1 \\ \frac{1}{2} [E\lambda_1 A + (1 - E\lambda_1)\theta] & \text{if } S = 0. \end{cases} \quad (4)$$

The expected increase in earnings is: $\Delta w = Ew(S = 1) - Ew(S = 0) = 0.5[g(A, \theta) - A]E\lambda_1$. Since $g(A, \theta) > A$, it follows that $\Delta w > 0$ for all A and θ .

If acquiring (cognitive) human capital were costless, then all individuals (with positive discount rates) would find it beneficial to attend school. Schooling is a costly investment and so, for the individual who invests in schooling,

$$\frac{1}{2}\beta [(A, \theta) - A] E\lambda_1 \geq C(A, \theta). \quad (5)$$

First, suppose that the costs of schooling do not vary with θ and are given by $c(A)$, with $c_A < 0$. In this case, there is only one value of A for which equation (5) holds with equality; this value is defined as A^* ; and it solves

$$\frac{1}{2}\beta [g(A^*) - A^*] E\lambda_1 = c(A^*). \quad (6a)$$

Since the left-hand side of equation (6a) is increasing in A^* and the right-hand side is decreasing in A^* , individuals for whom $A \geq A^*$ will choose to invest in schooling and the rest not. Hence, we should see sorting into different education levels according to the values of A .

Now suppose that the cost function and/or the discount rate β varies with θ . In this case, $A^*(\theta)$ is defined implicitly by

$$\frac{1}{2}\beta(\theta) [g(A^*(\theta), \theta) - A^*(\theta)] E\lambda_1 = C(A^*(\theta), \theta). \quad (6b)$$

Again, the left-hand side of equation (6b) is increasing in $A^*(\theta)$ and the right-hand side is decreasing in $A^*(\theta)$. Thus, for a given value of θ , all individuals for whom $A(\theta) \geq A^*(\theta)$ will choose to invest in schooling and all other individuals will choose not to do so. A^* varies with θ as follows:

$$\frac{dA^*}{d\theta} = -\frac{[\frac{1}{2}\beta_\theta(g-A) + \frac{1}{2}\beta g_\theta]E\lambda_1 - C_\theta}{\frac{1}{2}\beta(g_A-1)E\lambda_1 - C_A}. \quad (7)$$

The evidence in Castillo et al. (2011) suggests that misbehaving students have lower discount rates, which means that $\beta_\theta > 0$. By assumption, $g(A) > A$, $g_A > 1$, $g_\theta > 0$, $C_A \leq 0$, and $C_\theta \leq 0$. Therefore, $dA^*/d\theta < 0$. Observe that this result holds even if β is the only variable that depends on θ . Hence, the cutoff level of schooling investment, A^* , is decreasing in θ . As a result, conditional on A , we should see sorting into different education levels according to θ ; conditional on θ , we should see sorting into different education levels according to A . Note that in comparison to the previous case, in which neither the costs of schooling nor the discount rate varied with θ , here the individuals that are most affected by having low values of θ are those whose cognitive skills have values just above A^* defined by equation (6a). In the former case, they all would have chosen to invest in schooling, but in the latter case some decide not to invest in schooling. On the other hand, individuals with high noncognitive human capital may decide to acquire postsecondary education even if they have relatively low cognitive human capital (i.e., less than A^*).

We can now rewrite equation (4) as

$$Ew(S) = \frac{1}{2}E\lambda_1 A + \frac{1}{2}(1 - E\lambda_1)\theta + \frac{1}{2}(g(A, \theta) - A)SE\lambda_1, \quad (8)$$

where $S = 1$ if the individual decides to invest in schooling, and $S = 0$ otherwise. The implications of equation (8) for the earnings regressions are discussed in Section 2.3.

2.2. An Extended Model with Effort Choice

In the basic model, the relationship between cognitive and noncognitive skills follows from the assumption that noncognitive skills are an input in the production function of cognitive human capital. However, this is not the only way that such a relationship could be generated. In this section I endogenize that relationship by introducing effort into the basic model. In order to sharpen the result, I allow only the discount rate (and not the production and the cost-of-effort functions) to depend on θ . I therefore assume that $g(A, \theta)$ depends only on A (i.e., it is $g(A)$), and similarly that the costs of schooling are given by $c(A)$. However, I allow the production function of human capital to depend on (costly) effort, denoted by e . To simplify, I assume an additive production function in e (the results are qualitatively the same if a multiplicative function is used). Thus, the production function of cognitive human capital is now

given by

$$H_C = \begin{cases} g(A) + e & \text{if } S = 1 \\ A + e & \text{if } S = 0; \end{cases} \quad (9)$$

where the effort level $e \in [0, \bar{e}]$ and $g(A) > A + \bar{e}$. Thus, higher effort increases cognitive human capital but not more than schooling does. The costs of effort are given by $0.25Ke^2$, where K is a positive constant. The rest of the setting is the same as in the basic model. The worker determines his optimal effort level for each education level, then compares the costs to the benefits, and then makes his investment decision. The worker's first-stage problem is therefore

$$\max_{e \in [0, \bar{e}]} \begin{cases} \frac{\beta}{2} [E\lambda_1(g(A) + e) + (1 - E\lambda_1)\theta] - c(A) - \frac{1}{4}Ke^2 & \text{if } S = 1 \\ \frac{\beta}{2} [E\lambda_1(A + e) + (1 - E\lambda_1)\theta] - \frac{1}{4}Ke^2 & \text{if } S = 0 \end{cases} \quad (10)$$

Because of the additive form of the production function, the optimal effort choice is identical in both cases and is given by e^* :

$$e^* = \frac{\beta}{K}, \quad (11)$$

where e^* is the solution to the first-order conditions (assuming that $\beta \leq \bar{e}K$; otherwise, $e^* = \bar{e}$).¹³ The optimal effort level depends on θ via β , which means that e^* is increasing in θ .¹⁴

We substitute this effort choice into equation (9) and thereby obtain

$$H_C = \begin{cases} g(A) + e^* = g(A) + \frac{\beta}{K} & \text{if } S = 1 \\ A + e^* = A + \frac{\beta}{K} & \text{if } S = 0. \end{cases} \quad (12)$$

This equation implies that, even within the same educational category, cognitive human capital depends on θ through its dependence on β . Since $de^*/d\theta > 0$, cognitive human capital is increasing in θ .

After substituting for the optimal effort choice, we obtain that in the second period the worker's problem is

$$\max \left\{ \frac{\beta}{2} [E\lambda_1 A + (1 - E\lambda_1)\theta] + \frac{\beta^2}{4K}, \frac{\beta}{2} [E\lambda_1 g(A) + (1 - E\lambda_1)\theta] - c(A) + \frac{\beta^2}{4K} \right\}, \quad (13)$$

where the first term represents the choice of no schooling and the second an investment in schooling. Thus the individual will choose to invest in schooling if

13. The second-order condition is $-K < 0$, which guarantees that e^* is indeed a maximum.

14. If K depends on A and/or θ then it is reasonable to assume that the costs of effort are decreasing in A and/or θ (i.e., that $K_A < 0$ and $K_\theta < 0$). Under these assumptions, e^* is increasing in both A and θ .

$0.5\beta[g(A) - A]E\lambda_1 \geq c(A)$. Because β depends on θ , there is a cutoff level $A^*(\theta)$ such that, for a given value of θ , individuals for whom $A \geq A^*(\theta)$ will invest in schooling and all others will not. $A^*(\theta)$ is implicitly given by equation (6b) with g and c (i.e., the production and the cost-of-effort functions) that do not depend directly on θ . So substituting $g_\theta = 0$, $c_\theta = 0$, and $C_A = c_A$ in equation (7) yields

$$\frac{dA^*}{d\theta} = -\frac{\frac{1}{2}\beta_\theta(g - A)}{\frac{1}{2}\beta(g_A - 1)E\lambda_1 - c_A}, \quad (14)$$

which, under the assumptions, is negative, just as in the case of the baseline model.

Now the equivalent of equation (8) for this model can be written as follows:

$$Ew(S) = \frac{1}{2}E\lambda_1 \left(A + \frac{\beta}{K} \right) + \frac{1}{2}(1 - E\lambda_1)\theta + \frac{1}{2}(g(A) - A)SE\lambda_1. \quad (15a)$$

Note that if g is allowed to depend on θ , as in the baseline model, then the only change to the preceding equations will be to (14), where the numerator will now have an extra term: $0.5\beta g_\theta E\lambda_1$ that is positive.¹⁵ Thus, $dA^*/d\theta$ remains negative. In this case, equation (15a) becomes

$$Ew(S) = \frac{1}{2}E\lambda_1 \left(A + \frac{\beta}{K} \right) + \frac{1}{2}(1 - E\lambda_1)\theta + \frac{1}{2}(g(A, \theta) - A)SE\lambda_1. \quad (15b)$$

I discuss the implications of (15b) in the following section.

2.3. Implications and Relationship to the Empirical Specification

In translating the abstract model to an empirical specification, I associate eighth-grade misbehavior with low values of θ and associate high eighth-grade test scores with high values of A . Because skills are measured during the eighth grade, these fit best with interpreting the endowments θ and A in the model as corresponding to the human capital accumulated by that time. Since the eighth-grade test was designed to measure knowledge accumulated by this grade, it is natural to assume that the eighth-grade test scores measure A ; this interpretation suggests also that eighth-grade misbehavior could measure θ .¹⁶

By assumption, in both models, an increase in A leads to an increase in cognitive human capital. The interpretation is straightforward: those who are endowed with (or have accumulated) a higher current stock of knowledge, will naturally end up with greater knowledge.

15. Since $g_\theta > 0$, in this case the effort is increasing in θ for the highly educated group, too.

16. This interpretation suggests that the eighth-grade measures are enough to describe the eighth-grade stock of both types of human capital. It does *not* mean that, in the process of producing each type of human capital, the other type was not used. Because I have no data to test for the nature of that production process, this assumption seems best.

The situation with noncognitive human capital is more subtle. The baseline model postulates a positive relation between θ and cognitive human capital, which suggests that boys who do not do their school work know less. In the extended model, this is a result not an assumption. In particular, individuals with higher values of β are willing to pay the costs of effort and to behave well (in eighth grade) in order to reap the future benefits of higher cognitive human capital (and possibly higher education). Conversely, students who misbehave could easily have the faculties needed to accumulate as much knowledge as their fellow students who behave well, yet misbehaving students – because of their their low discount rates – choose to put in less effort and hence will have lower levels of cognitive human capital (which may be manifested by lower tenth-grade test scores).¹⁷

The most straightforward interpretation of the schooling decision in the models is whether to obtain a postsecondary degree. As long as β is related to θ , both models imply that only individuals with sufficiently high endowment of both cognitive and noncognitive human capital will invest in schooling. The models suggest that the probability of obtaining a postsecondary degree is decreasing in misbehavior and increasing in test scores. Moreover, individuals with (relatively) high cognitive human capital will be the ones most affected by misbehavior. Even though their cognitive skills are high enough to attend college, these individuals will decide against it because they cannot bear the costs associated with good behavior and/or they discount the future to such an extent that the costs, acquired in the present, outweigh the future benefits. In the educational attainment regression we therefore expect to find, for high education levels, a negative sign on the term for the interaction between the measures of cognitive and noncognitive human capital.

Equation (8) and equation (15b) each imply that, conditional on cognitive ability, those who misbehaved in eighth grade have lower earnings on average. Conditional on eighth-grade misbehavior, those who had higher eighth-grade test scores have higher earnings on average. Since $g(A, \theta) - A > 0$, the last term in these equations (i.e., $0.5[g(A, \theta) - A]SE\lambda_1$) means that the correlation between cognitive human capital and earnings is higher for highly-educated than for low-educated workers.¹⁸

17. There is direct evidence of positive correlations between test scores and impatience (e.g., Benjamin et al., forthcoming; Dohmen et al., 2011).

18. A similar relationship holds even in the most general case. In that case, we write the cognitive human capital of workers with $S = 0$ as $g(A, \theta, e_{S=0}^*)$ and of workers with $S = 1$ as $g(f(A), \theta, e_{S=1}^*)$, where $f(A) > A$ for all A and where $e_{S=i}^*$ is the optimal effort level of a worker with schooling i . Here, the equivalent to equation (15b), which we refer to as (15b*) is $Ew(S) = 0.5\{E\lambda_1 g(A, \theta, e_{S=0}^*) + (1 - E\lambda_1)\theta + [g(f(A), \theta, e_{S=1}^*) - g(A, \theta, e_{S=0}^*)]SE\lambda_1\}$. Since those who invest in schooling should have higher cognitive human capital, it follows that $g(f(A), \theta, \bar{e}) > g(A, \theta, \bar{e})$; this implies $g(f(A), \theta, e_{S=1}^*) - g(A, \theta, e_{S=0}^*) > 0$. Therefore, the last term in (15b*) is increasing in A , and thus highly-educated workers experience higher returns to cognitive human capital than low-educated ones.

Observe that if the worker's productivity depends on his contemporaneous effort choice whereas his expected rewards are in the future (e.g., a greater likelihood of keeping the job or receiving of promotion) then those with lower values of β (i.e., lower values of θ) will invest less effort (i.e., shirk), which could explain why firms would directly care about this specific noncognitive skill. This argument is similar to the one in Bowles et al. (2001b).

Further implications depend on functional forms and modeling assumptions. For example, if $g(A, \theta)$ is nonseparable in A and θ then the model suggests that there will be an interaction term between the measures of cognitive and noncognitive human capital for individuals who choose to acquire schooling. This is also the case if the production function is not additive in effort and A . If K depends on A then we should also expect, through the term β/K , an interaction term between the measures of cognitive and noncognitive human capital for workers with $S = 0$.¹⁹ While one can derive specific predictions given other functional forms for g , these will depend strongly on functional forms and modeling assumptions. For example, for the models described above, if g is additive in A and θ or if g does not depend on θ then there will be no interaction terms between cognitive and noncognitive human capital. However, if the production function of cognitive human capital is multiplicative in effort – so that $H_C = g(A, \theta)e$ if $S = 1$ and $H_C = Ae$ otherwise – there will always be an interaction term between cognitive and noncognitive human capital. In Section 4 I will therefore discuss the functional form of g and the models that are supported by the data described in Section 3.

3. Data

The data set used for the analysis is the National Educational Longitudinal Survey (NELS) conducted by the National Center for Education Statistics. A nationally representative sample of eighth-graders was first surveyed in 1988, after which a subsample of the respondents was surveyed again with four follow-ups in 1990, 1992,

19. Such an interaction term may exist also when $K_A = 0$ if the basic model is extended to allow cognitive human capital to depend on θ even for workers with $S = 0$ (with the interpretation that, between eighth grade and when postsecondary schooling decisions are made, individuals may have acquired more cognitive human capital and that this process may depend on θ). Equation (9) thus becomes $H_C = g(A, \theta) + e$ if $S = 0$ and $H_C = g(f(A), \theta) + e$ if $S = 1$. Since those who invest in schooling should have higher cognitive human capital, we need that $f(A) > A$ and $g(f(A), \theta) > g(A, \theta)$ for all A and θ . To ensure this, we can assume that $g_A > 0$ and $f'(A) > 1$ or that $g_{AA} > 0$ and $g_{A\theta} > 0$; the only change is to equations (14) and (15b). Under the assumptions, $dA^*/d\theta$ remains negative and (15b) becomes $Ew(S) = 0.5\{[g(A, \theta) + \beta/K]E\lambda_1 + \theta(1 - E\lambda_1) + [g(f(A), \theta) - g(A, \theta)]SE\lambda_1\}$. This equation implies that there should be an interaction term between cognitive and noncognitive human capital also for low-educated individuals when g is non-separable in A and θ .

1994, and 2000. The study provides detailed family background data as well as information about participants' activities before, during, and after high school. The 1994 and 2000 follow-ups contain detailed information on high school graduation status, postsecondary education, and labor market outcomes. A test battery was administered in the base year and in the first two follow-ups. In addition, students' teachers, parents, and school administrators were surveyed. Teacher evaluations of student behavior were used to construct the behavioral variables, which are detailed in the subsequent text.

Outcomes examined in the analysis include respondents' test scores in tenth grade (based on data in the 1990 survey), educational attainments as of 2000 (from data in the 1994 and 2000 surveys), and labor market outcomes as of 1999 (from data in the 2000 survey). In addition, parental characteristics (taken from the 1988 parent survey) are used as controls. The construction of these variables and their respective means are described in Table WA2 of Web Appendix A.

3.1. Behavioral Measures

The measures of behavior used in the analysis are based on eighth-grade teacher evaluations of student behavior in the classroom.²⁰ Students were evaluated by two of their teachers; the first taught math or science, and the second taught English or social science. The evaluations consist of yes/no answers to the following questions. Is [the student] frequently absent? Is [the student] frequently tardy? Is [the student] consistently inattentive in class? Is [the student] frequently disruptive? Does [the student] rarely complete his homework?

The behavioral measures were coded such that low values correspond to good behavior and high values to misbehavior. I use two measures in the analysis. In the first, each of the five evaluations was coded as a dummy variable that takes the value 1 if at least one teacher indicated that the student misbehaved on the category and is set to 0 otherwise. The means and the conditional means of these variables are reported in (respectively) Tables WA1a and WA1b of Web Appendix A. For the second, the first principal component was used to construct a composite behavioral measure consisting of the weighted sum of the ten teacher evaluations (five per teacher);²¹ this is the "misbehavior" variable. This variable was normalized to have weighted mean 0 and

20. The NELS also includes teacher evaluations for the tenth and twelfth grades. However, these evaluations supply information on a smaller set of students (because of decreasing sample size by design) that represent a selected sample (because some students participating in the NELS dropped out of school after eighth grade). Furthermore, only a subset of teachers were asked to evaluate their twelfth-grade students. It is for these reasons that I use only the eighth-grade teacher evaluations in the analysis.

21. The weights are as follows: Tardiness: 0.26 and 0.27 for T1 and T2 (i.e., for teacher 1 and teacher 2, respectively); Absenteeism: 0.25 and 0.26 for T1 and T2; Disruptiveness: 0.30 for both

standard deviation of 1, for the sample of 9,335 male eighth-graders who in this grade had two teacher evaluations as well as valid test scores. I use both the different behavioral traits and the composite variable as explanatory variables in the analysis.²²

Reliability of Teacher Evaluations. There is evidence in the literature suggesting that teacher race and gender may bias their evaluation of some students (e.g., Dee, 2007; Lavy, 2008). However, the studies that use the NELS and look at additional outcomes (Ehrenberg et al., 1995; Dee, 2007) find little evidence for biased teacher evaluations of boys.²³ So the teacher evaluations used here, besides being comprehensive and helping to avoid self-reporting problems, seem also to be unbiased in this sense.

A different question is whether teacher evaluations of students' behavior are related to the students' noncognitive traits. The evidence in Castillo et al. (2011) reveals a relation between discount rates and disciplinary referrals. Therefore, as long as teacher evaluations reflect their records of how students behaved in the classroom, one can reasonably suppose that teacher evaluations are indeed related to their students' noncognitive traits. The evidence in psychology leads to similar conclusions. Barbaranelli et al. (2003) demonstrate that teacher evaluations of their junior high students matched similar reports made by the students and their mothers regarding the students' personality traits and behavior. These authors show that students' personality traits (as assessed by their teacher) are correlated with the behavior of these students (as reported by the teachers themselves, the students, and their mothers). As mentioned in the introduction, Barbaranelli et al. (2003) also document that low conscientiousness is correlated with both externalizing behavior (including misbehavior in school) as well as low grades in school.

3.2. Test Scores

The NELS participants completed three sets of cognitive tests in each of the first three survey rounds. Since a substantial number of the twelfth-grade scores were

teachers; HW: 0.37 and 0.38 for T1 and T2; Inattentiveness: 0.36 and 0.37 for T1 and T2. In all cases T1 denotes the math or science teacher and T2 the English or social science teacher.

22. The reported results remain qualitatively the same when I use a less detailed measure (used in Figure 1 and Table 1) – namely, an indicator variable set equal to 1 if *any* teacher reported misbehavior in any category. Additionally, when I add to the regressions that use the five misbehavior measures another five dummy variables, each set equal to 1 only if *both* teachers reported misbehavior in the category, these dummies are never jointly significant in the regressions.

23. Ehrenberg et al. (1995) find evidence that teacher race and gender are related to their evaluation of girls only. While Dee (2007) finds that having a female teacher is negatively related to eighth-grade teacher evaluations of boys' disruptiveness and homework completion, he also finds that having a female teacher is negatively related to eighth-grade achievements of boys and decreases boys' engagement with the teacher subject. This suggests that having a female teacher actually affects boys' behavior in school and that the evaluations themselves are not biased.

imputed, only the eighth- and tenth-grade test scores are used in analysis. Four different subtests were developed for the NELS; they covered reading comprehension, mathematics, science, and history/citizenship/geography. In addition, the NELS data set contains a composite mathematics and reading test score. The tests were designed to check whether students had the knowledge they were supposed to have acquired by a certain grade. Specifically, the first (second) test battery, administered in 1988 (1990), was designed to test knowledge accumulated by the eighth (tenth) grade.

Studies that investigate the relationship between test scores and labor market outcomes often use composite math and reading scores. To enable comparison between this paper's results and those in the literature, I use the NELS-constructed composite math and reading test scores. These composite scores were normalized to have weighted mean 0 and standard deviation of 1 for the sample of 9,335 male eighth-graders who had in this grade two teacher evaluations and valid test scores.

In Section 5.2, I use the tenth-grade composite math and reading test scores as a proxy for cognitive human capital accumulated by the tenth grade. These scores were likewise normalized to have weighted mean 0 and standard deviation 1 for the 5,161 male tenth-graders who in this grade had two teacher evaluations and valid test scores.²⁴ In eighth grade, all participants were given the same test; in tenth grade, however, multiple test forms were used. The purpose was to target potential varying knowledge levels in the student population. Two forms of the reading test were developed, and students who scored (in eighth grade) above the mean were given the more difficult form. For the mathematics test, three forms were developed. The easiest and the hardest of these forms were administered to students who had scored in the lowest and highest quartiles, respectively, in eighth grade; students in the middle half of the distribution received the test of middle difficulty. The NELS administrators employed item response theory (IRT) to facilitate comparisons between participants who took different tests.

3.3. NELS Design Problems and Weights

Several features of the NELS sample design are crucial for the analysis (for an excellent summary, see Grogger and Neal, 2000). In the 1990, 1994 and 2000 follow-ups, only a subsample of NELS participants was retained. The likelihood of a particular

24. Students who were not in tenth grade were excluded for the following reasons. Whereas more than 95% of the in-school participants completed the test batteries, only about 50% of the dropouts did so. In order to avoid selection issues in the dropout pool, only the test scores of the in-school students were used. Moreover, the 1990 test battery was designed to measure knowledge acquired by the tenth grade; by definition, ninth-grade students or dropouts (who did not learn the material) have less knowledge. Excluding these students prevents them from driving the results – although including these out-of-grade students does not qualitatively change the results reported in Section 5.2.

student participating in any given survey was not determined at random. Instead, it was determined by his dropout status and the quality of his response to past questionnaires as well as his ethnicity/race and socioeconomic status. Because of these nonrandom retention procedures, the resulting NELS sample is unrepresentative of the US population – even within gender and race/ethnicity cells. Hence, the sample weights supplied with the data set are used throughout the analysis. Specifically, for the outcomes analysis I use the weights that account for the sampling procedures used from the 1988 survey through the 2000 survey. For the analysis in Section 5.2, I use the weights that account for the sampling procedures used in the 1988 and 1990 surveys.

4. Outcomes

In this section, I present evidence of a meaningful economic association between behavior in eighth grade and both educational and labor market outcomes for young men. Table 1 presents the basic characteristics of the outcome variables of interest. The table depicts the means of these variables, and then breaks them down in terms of an eighth-grade misbehavior dummy set equal to 0 for all individuals whose eighth-grade teachers both reported good behavior in all five categories (and set to 1 otherwise). By this division, about 55% of male eighth-graders misbehaved.

Even the simple means convey a clear picture. For example, those who misbehaved in eighth grade are almost 3 times more likely to drop out of high school and almost 3 times less likely to graduate from college. Eleven years later, those who misbehaved in eighth grade earn on average 9.8% less per year and 8.5% less per hour than those who behaved well. Using the binary indicator for misbehavior, I plot in Figure 1 the kernel densities of earnings, hourly wages, and total hours worked in 1999 for the two groups. The figure clearly indicates that the earnings and hourly wages distributions of misbehaving students both lie to the left of the corresponding distribution for well-behaved students. This supports the notion that individuals who misbehave in eighth grade have worse labor market outcomes later in life. Because the two distributions of total hours worked are very similar, the differences in the outcomes are not the result of leisure–work substitution.

The simple statistics just presented do not take other factors into account. In particular, one cannot infer thereby whether eighth-grade misbehavior is associated with educational outcomes once test scores and parental inputs are taken into account. Nor do we know whether misbehavior is associated with labor market outcomes after test scores and educational attainment are accounted for. Below I investigate these questions in detail.

Each of the empirical specifications below takes the form

$$Outcome_i = f(TS_i, MB_i, X_i, \varepsilon_i),$$

where i indexes individuals, *Outcome* is the outcome of interest, *TS* denotes eighth-grade test scores and *MB* eighth-grade misbehavior, X is a vector of other covariates, and ε is an error term.

4.1. Educational Attainment as of 2000

I start by investigating the relationship between misbehavior in eighth grade and educational attainment. The simple means in Table 1 suggest that misbehavior is negatively correlated with both test scores and educational attainment. Further descriptive evidence can be found in Figure 2, which depicts the cumulative distributions of eighth-grade test scores and misbehavior by schooling levels as of 2000. The figure plots sorting by different education levels according to eighth-grade test scores and the composite misbehavior variable. The left panel depicts a well-known phenomenon: people who have higher test scores tend to acquire higher levels of schooling. The right panel reveals a less known phenomenon: those who behave better tend to have higher levels of schooling.

The models make predictions regarding the decision to acquire postsecondary education, so I investigate this issue here. Thus I report the results from an empirical (probit) model predicting the likelihood of acquiring postsecondary education – that is, at least, an Associate of Arts (AA) degree. In Web Appendix B, I analyze in detail the results of using a model that takes more educational choices into account.

The first two columns of Table 2 report the marginal effects, evaluated at the sample means, of the probit regression just described. In the first column, the composite misbehavior variable is used; in the second column the five misbehavior measures are used. Focusing on the first column, we note that eighth-grade misbehavior is significantly and negatively associated with the probability of obtaining a postsecondary degree.²⁵ Thus, evaluated at the sample means, an increase of 1 standard deviation in eighth-grade misbehavior is associated with an 8.6% decrease in the probability of graduating from college or obtaining an AA degree. This number is of the same order of magnitude as the association between eighth-grade test scores and acquiring a postsecondary degree. Since the predicted probability of obtaining a postsecondary degree is itself 31.1%, the 1-standard deviation (thereafter, 1-SD) increase in misbehavior represents a decrease of about 28% in this probability. When

25. The results presented in Web Appendix B confirm that the negative correlation is indeed with obtaining a postsecondary education – in particular, a bachelor’s degree – and not with the likelihood of graduating from high school.

the five measures of misbehavior are used (in column 2) instead of the composite variable, the findings are similar. Even though the coefficients for each of the five measures are not always statistically significant, jointly they are significant. Changing one from being one of the best-behaved eighth-graders (i.e., all five indicators of misbehavior are 0) to being one of the worst-behaved eighth-graders (i.e., all indicators are 1), is related to a 21.8% decrease in the likelihood of obtaining a postsecondary degree.

The models suggest that individuals with relatively high knowledge are most affected by misbehavior. As explained in Section 2.3, such students may forgo college despite having sufficient cognitive skills to pursue a higher education because they have “too” low noncognitive skills. We thus expect to see a negative sign for the interaction term between the measures of cognitive and noncognitive human capital in the educational attainment regressions. Column 3 of Table 2 gives the marginal effects at the sample means when this interaction term is added to the specification of column 1. The interaction term is significant and negative, as predicted by the models. In particular, a 1-SD decrease in misbehavior is associated with a $9.1\% + 3.8\% \times (\text{test scores})$ increase in the likelihood of obtaining a postsecondary degree. Thus, individuals who behave well benefit the most (in terms of the probability of obtaining a postsecondary degree) from having high test scores and, conversely, those who misbehave benefit the least.

The relationship between family background characteristics depict in Table 2 are in accordance with the literature (see Cameron and Heckman, 2001; Jacob, 2002).²⁶ Higher parental income is positively related to higher levels of education, as is higher parental education.²⁷ The table suggests that not living with both parents is associated with a nearly 13.5% decrease in the predicted probability of obtaining a postsecondary degree. After family background is controlled for, there seems to be no gap in postsecondary educational attainment between blacks and whites.²⁸ Although these results may seem surprising, they are consistent with the literature. Jacob (2002) reports similar findings for the NLES; Cameron and Heckman (2001) find that after

26. Family background characteristics are included in the regressions because they might provide information on credit constraints faced by individuals. The values of these variables are taken from the 1988 survey (when participants were in eighth grade). For the construction of the variables, see Table WA2 in Web Appendix A.

27. There are some differences between the Table 2 and the results presented in Web Appendix B – especially in the effect of parental education. These differences are the result of a coarser division of educational choices in Table 2.

28. Here, too, the results described in Web Appendix B are somewhat different. Specifically: when a finer division of education level is used, being black is associated with an even lower probability of *not* graduating from high school and being Hispanic with an even higher probability of attending some college. The difference is again explained by the merging together of several education levels in the regressions of Table 2.

they control for family background characteristics in the 1979 National Longitudinal Study of Youth (NLSY), minorities are more likely than whites to graduate from high school and attend college.

The findings reported in this paper indicate that eighth-grade misbehavior is associated with lower educational attainment (once eighth-grade test scores and family characteristics are controlled for). In additional regressions presented in Web Appendix B that feature a finer division of educational attainment, I find that eighth-grade misbehavior has the strongest relationship with the probabilities of graduating from college and of dropping out of high school, yet has weaker relationship with the probability of obtaining other levels of schooling. The positive relationship between dropping out of high school and misbehavior is to be expected. The negative relationship between misbehavior and the probability of graduating from college suggests that misbehavior may be indicative of more than a temporarily defiant phase in one's life.

4.2. Labor Market Outcomes: Earnings in 1999

Given the known relationship between higher education levels and labor market success, we expect to find a negative correlation between misbehavior in eighth grade and labor market outcomes. Indeed, the raw statistics support this conjecture. In this section, I investigate in detail the relationship between eighth-grade misbehavior and earnings.

Table 3 presents the earnings regressions. The dependent variable is the ln of earnings in 1999.²⁹ The sample was restricted to include only males for whom eighth-grade test scores and two teacher evaluations were available. It was restricted further to include only individuals who were not enrolled in a postsecondary institution after January 1999,³⁰ were full-time workers, were not self-employed and reported positive income in 1999.³¹ Columns 1 and 2 report the coefficients from the earnings regressions, which control for (respectively) demographics and eighth-grade test

29. Since the NELS participants were asked about their earnings in 1999 and not about their hourly wage, I use the variable with an a priori smaller measurement error. The regression results in which the dependent variable is the ln of hourly wage are both quantitatively and qualitatively similar and can be found in my PhD dissertation (Segal, 2005).

30. Of all the male participants, 29% reported being in school after 1999. On average this group behaved better in eighth grade and had higher eighth-grade test scores (the difference in the means – of the eighth-grade misbehavior composite variable and the eighth-grade test scores – between the two groups are 0.1 and 0.4 standard deviations, respectively).

31. Full-time workers are those who reported working at least one week in 1999 and an average of at least 35 hours per week during the period worked. Of the 2,323 men who were not enrolled in school and for whom valid test scores, behavior, and employment data are available, 2,240 (96.4%) worked full-time. The regression results for all workers (i.e., with positive amount of hours) are very similar qualitatively and quantitatively and can be found in Segal (2005).

scores. These regressions document that, for the NELS data set as well, minorities earn significantly less; once test scores are controlled for, however, these differences are no longer significant. Columns 3 and 4 add to the regression of column 2 the eighth-grade misbehavior variables. Misbehavior in eighth grade is clearly associated with a significant reduction in earnings controlling for test scores: a 1-SD increase in misbehavior (i.e., worse behavior) is related to a 6% decrease in earnings. Moreover, the association between misbehavior and earnings has nearly the same magnitude as the association between test scores and earnings: a 1-SD decrease in eighth-grade test scores is associated with a 5.4% decrease in earnings. The results are similar when I use the five misbehavior measures. Even though the coefficients on each of the five measures are not always significant by themselves, jointly they are statistically significant.³² The change from being one of the best-behaved (i.e., all five misbehavior indicators are 0) to one of the worst-behaved eighth-graders (i.e., all indicators are 1), is associated with a 20.6% decrease in earnings.

In Columns 5 and 6, years of schooling completed by 1999 are added to the regressions. Comparing column 5 to column 3 shows that after years of schooling completed are controlled for, the coefficient for misbehavior is reduced in size – although it remains statistically significant and economically large: a 1-SD increase in misbehavior is associated with a 4% decrease in earnings.³³ Column 6 uses the five misbehavior measures whose coefficients are again jointly significant.³⁴ The change from being one of the worst-behaved to one of the best-behaved eighth-graders, *ceteris paribus*, is related to a 18.4% increase in earnings. These decreased magnitudes indicate that the relation between misbehavior and labor market outcomes is partly mediated by the relation between misbehavior and educational attainment.

Once education levels are controlled for, neither statistically nor economically significant associations between eighth-grade test scores and earnings are found in the data (columns 5 and 6 in Table 3). This result does not stem from adding the misbehavior measures to the regressions. After years of schooling completed are added to column 2's regression, the coefficient for eighth-grade test scores is small (0.023) and insignificant. The last three columns of Table 3 address this issue. In column 7, the coefficients for test score are allowed to vary between those who obtain a

32. The coefficient for disruptiveness in column 2 is positive. This coefficient is negative, however, when disruptiveness is the *only* measure of behavior, suggesting that the positive sign is due to correlations between the five measures.

33. The change in statistical significance is not the result of less precise estimation, as the standard errors of the misbehavior coefficients hardly change. Rather, it is the point estimate that is being reduced in magnitude by about a third.

34. Comparing columns 6 and 4, we observe that the traits that have the strongest association with educational attainment (i.e., homework completion and absenteeism) are reduced in magnitude the most when education is controlled for.

postsecondary degree (a category that includes the AA degree) and those who at most had just some college education. The figures clearly indicate that higher eighth-grade test scores are associated with higher earnings only for those males who obtain postsecondary degrees.³⁵ It seems that once education is controlled for, test scores do not correlate with earnings for those who have no more than “some” college education.³⁶ Columns 8 and 9 add the different eighth-grade misbehavior measures to the regressions. The coefficient for misbehavior is large and statistically significant: a 1-SD increase in misbehavior is associated with a 4.3% decrease in earnings. The five measures are, again, jointly significant, and the increase in earnings associated with turning one from being one of the worst-behaved to one of the best-behaved eighth-graders is 19.1%.³⁷

The regressions reported in the last three columns of Table 3 are in accordance with the models’ predictions that associations between test scores and earnings will be larger for those who obtain a postsecondary degree. The models yield other predictions that are not tested in Table 3. Specifically, depending on the functional form of the cognitive human capital production function, we may expect the associations between misbehavior (i.e., noncognitive human capital) and earnings to depend on the schooling level and may also expect an interaction term between misbehavior and test scores in the earnings regressions. Below, I use the composite misbehavior variable to test for these predictions.

The first column of Table 4 tests for the most general form of cognitive human capital production function. Thus, it includes interaction terms between eighth-grade test scores and misbehavior that vary by education levels, and it allows both test scores and misbehavior to vary by education level. In the table’s second column, the interaction term is constrained to be the same across education levels; in the third column, the interaction term relates to the earnings of highly educated men only

35. The difference between the two coefficients for test scores is significant ($p = 0.0014$), and jointly they are significantly different from zero ($p = 0.0013$).

36. Segal (forthcoming) reports similar findings for male NLSY participants. Specifically, she finds that the association between Armed Forces Qualification Test (AFQT) scores and earnings in 2003 is twice as strong for men who graduated from college as for men with lower educational attainment.

37. So far I have treated the misbehavior measures as absolute measures, yet it is possible that the teachers answered questions on student behavior in a relative manner. In Table WC1 of Web Appendix C, I investigate this issue by taking advantage of the NELS sample’s inclusion of multiple students from each school. I restrict the sample further to include only those schools in which at least two students had valid data, so about 10% of the sample used in Table 3 is discarded. The first four columns of Table WC1 present the basic earnings regressions for the restricted sample, and the last four columns present the regression with eighth-grade school fixed effects. Adding school fixed effects to the earnings regressions does not change the main findings on the association between eighth-grade misbehavior and earnings. Moreover, the regression results show that eighth-grade schools are related both to educational attainment and to eighth-grade test scores. This finding may serve as indirect evidence for the importance of school quality or of sorting of students to eighth-grade schools.

(as predicted by the models). The results in Table 4 confirm that test scores are correlated only with earnings of highly educated men. In contrast, the associations between misbehavior and earnings are practically the same for all men (although the coefficients are less precisely estimated for the highly educated). Moreover, the first three columns suggest that there is no interaction term between cognitive and noncognitive human capital. Yet because this lack of interaction could result from (erroneously) specifying that misbehavior varies across education levels, columns 4 to 6 repeat the regressions with misbehavior constrained to be the same across education levels. The results remain unchanged – in other words, there is no significant interaction term between eighth-grade test scores and misbehavior in the earnings’ regressions. Column 7 confirms that indeed the associations between misbehavior and earnings are the same across education levels; it also suggests that the specification of column 8 of Table 3 is a correct one, since the coefficients for eighth-grade misbehavior are literally the same across education levels. Thus, the preferred specifications for the rest of this paper are the two presented in columns 8 and 9 of Table 3.

Taken together, the results presented in Table 4 – specifically, the lack of interaction term – are consistent with models in which the production function of cognitive human capital does not depend directly on noncognitive human capital. These results are also consistent with a cognitive human capital production function that is additive in effort and with a cost-of-effort function that does not depend on cognitive human capital. Although one cannot rule out the possibility that noncognitive human capital (related to misbehavior) is a direct input in the production function of cognitive human capital, the results in Table 4 put limitations on this possibility. In particular, the lack of differential associations between eighth-grade misbehavior and earnings across education levels suggests the following. If misbehavior is indeed a direct input in the production function of cognitive human capital, then it must enter not only additively but also it has the same functional form for cognitive human capital produced in high school and in postsecondary education.³⁸

5. Channels

This section explores the possible explanations for the associations between misbehavior and economic outcomes. First I consider economic explanations that

38. To see this note first that the last term in equation (15b) suggests that if $g(A, \theta) = g(A) + h(\theta)$ then θ should be more strongly correlated with earnings of highly educated workers. Hence in order for θ to be an input in the production function and for the results of Table 4 to hold, we must be in the world described in footnote 19. Thus, in this case, the last term of (15b) is $g(f(A), \theta) - g(A, \theta)$; for g to be additive in A and θ , it must be that $g(f(A), \theta) = g(f(A)) + h(\theta)$ and $g(A, \theta) = g(A) + h(\theta)$.

imply that these relationships do not stem from eighth-grade misbehavior causing worse economic outcomes. I then provide evidence that misbehaving eighth-graders have worse tenth-grade test scores. However, neither a boy's background nor his above-average knowledge in eighth grade can alleviate this effect. Taken together this suggests boys who misbehaved in eighth grade most likely kept misbehaving and not only that misbehavior results in less accumulated cognitive human capital. It thus seems that misbehavior in eighth grade is an indication of one's type, and it is this type that (to a large extent) determines the individual's behavior in school and the workplace.

5.1. Taste for Occupation or Education

In this section, I investigate whether misbehavior is merely a by-product of an omitted variable that itself is causally related to outcomes. Below, I investigate whether tastes and expectations are these omitted variables. The mechanism of this channel is similar to the one discussed in the extended model. Specifically, suppose that eighth-graders have different tastes for different occupations, irrespective of their skills. Schooling and/or knowledge acquired in school is not a prerequisite for some of these occupations, so eighth-graders who aspire only to such jobs may decide not to invest effort in school. Similarly, suppose that eighth-graders have different tastes for education, regardless of their skills; those with no taste for schooling may decide not to invest effort in school. Less effort will result in less knowledge, which (all else being equal) leads to lower future earnings. The NELS eighth-grade respondents were asked: "What kind of work do you expect to be doing when you are 30 years old?" They were also asked: "As things stand now, how far in school do you think you will get?" I can therefore test these two hypotheses directly. The regression results are presented in Table 5.

The first two columns of Table 5 report the results from the preferred specifications (i.e., of columns 8 and 9 of Table 3) for a restricted sample that includes only eighth-graders who reported their expected occupation at age 30. In columns 3 and 4, dummies for each category of expected occupation are added. The coefficients for eighth-grade misbehavior are essentially the same, which means that there is no evidence that the association between eighth-grade misbehavior and earnings results from taste differences with respect to future occupations.³⁹

Columns 5 and 6 report results from the preferred specifications for a restricted sample that includes only eighth-graders who reported their expected education at

39. In specifications where dummies for the categories "not working", "do not know", and "other" were added, the results were essentially the same.

age 30,⁴⁰ and the last two columns of Table 5 add the variable “expected years of schooling” to the regressions. The coefficients for eighth-grade misbehavior are essentially the same, so there is likewise no evidence that the association between eighth-grade misbehavior and earnings stem from differences in tastes for education.

If test scores measure one’s cognitive ability, then the results reported in Table 5 directly rule out the following channel: eighth-graders with low cognitive ability realize that the occupation in which they could succeed as adults are those requiring only low levels of human capital (or realize that they could not succeed much further in school); hence they decide to misbehave in school. If this were the case, then adding eighth-grade test scores to the earnings regressions would have rendered eighth-grade misbehavior insignificant.

5.2. *Cognitive Human Capital Accumulation*

Another possible explanation is that misbehavior in school interferes with cognitive human capital accumulation. All inputs in the production function of human capital should affect test scores. The inputs that received the most attention are cognitive ability, school quality, and (to the extent it is not captured by those) parental characteristics. In addition to these factors, all educators agree that doing one’s schoolwork contributes to one’s knowledge. Thus, students who misbehave in eighth grade may learn less and, as a result, will have lower cognitive human capital. The NELS data set contains two test batteries designed to measure knowledge that students are supposed to have accumulated by eighth and tenth grades. I use these test scores as measures of cognitive human capital in order to examine directly the relationship between eighth-grade misbehavior and knowledge accumulation by the tenth grade. In addition, I allow tenth-grade test scores to be affected by school quality (captured by eighth-grade school dummies in the regressions) and eighth-grade parental characteristics. Thus, I estimate the model: $TS_{ij}^{10} = \alpha_j + X_{ij}\beta + \beta_{MB}MB_{ij} + \beta_{TS}TS_{ij}^8 + \varepsilon_{ij}$, where i indexes individuals and j indexes the eighth-grade schools. Here the dependent variable TS^{10} denotes the tenth-grade standardized test scores, the α_j are the school specific coefficients, X is a vector of demographics and family characteristics; TS^8 denotes the eighth-grade standardized test scores, MB denotes measures of eighth-grade misbehavior, and ε is an error term.

40. The answers were translated to expected years of schooling via this coding procedure: “Won’t finish high school” was coded as 11 years of schooling; “Will finish high school or attend vocational, trade, or business school after high school” was coded as 12 years of schooling; “Will attend college” was coded as 13 years of schooling; “Will finish college” was coded as 16 years of schooling, and “Higher schooling after college” was coded as 18 years of schooling. In parallel specifications where dummies were included for each possible answer, the results were both qualitatively and quantitatively the same.

Table 6 reports the results. The first three columns omit the eighth-grade test scores from the regressions. Column 1 includes only family characteristics and eighth-grade school dummies; in columns 2 and 3, the misbehavior measures are added. Column 1 reveals the expected relationship between family characteristics and tenth-grade test scores. Columns 2 and 3 suggest that part of that relationship is mediated by eighth-grade misbehavior, since the point estimates on family characteristics are decreasing once eighth-grade misbehavior is added to the regressions. Eighth-grade misbehavior is associated with lower tenth-grade test scores. This relationship is of the same order of magnitude as the relationship between family background characteristics and tenth-grade test scores and is statistically significant (again, the five measures of eighth-grade misbehavior are jointly significant). Columns 4 and 5 add the eighth-grade test scores. The results indicate that eighth-grade misbehavior still correlates with tenth-grade test scores. Both the composite misbehavior variable and the five misbehavior measures are statistically significant, but only the race/ethnicity dummies and the family composition dummy remain significant and large. One can thus infer that the correlations of eighth-grade misbehavior with educational attainment and with earnings may be due in part to less human capital accumulated by the students who misbehaved in eighth grade.⁴¹

The evidence suggests that misbehavior results in lower cognitive human capital. However, in order to understand the relationship between misbehavior and earnings, we must establish whether or not misbehavior in eighth grade alone – that is, without any further incidents of misbehavior – is enough to account (via continued loss of cognitive human capital) for the lower earnings observed for misbehaving boys later in their lives. If it is enough, then a boy who experienced a negative shock in eighth grade will have lower tenth-grade test scores even if he behaved extremely well in ninth grade.⁴² Therefore, in the rest of this section I investigate whether the relationship depicted in the first five columns of Table 6 can be the result of the boys' misbehavior

41. Not all tenth-graders took the same test, so it is possible that the relationship between eighth-grade misbehavior and tenth-grade test scores reflects this fact. In particular, it may be that the IRT procedure, applied by the NELS administrators, assigned a too-low compatible score to individuals with low eighth-grade scores and that this creates the associations found in the data. If that were the case, then misbehavior may not actually affect human capital accumulation (which in turn suggests that this channel cannot explain the relationship found in the data). When the regressions in column 4 are repeated by subgroups that took the same tests in tenth grade, I find that eighth-grade misbehavior is significantly associated with tenth-grade scores for all groups other than the worst one (i.e., students in the lowest eighth-grade quartile of math scores and with below-average reading scores).

42. Examples of such shocks are illness or the death of a relative while in eighth grade, but also a specific eighth-grade taste for occupation, aversion to school, or erroneous expectation regarding the future.

only in eighth grade.⁴³ If the eighth-grade misbehavior was an isolated incident then one would expect to find two other relationships in the data. First, families with more resources may be able to help their children “catch up.” Thus, higher family income or highly educated parents might have a positive influence on tenth-grade test scores even after eighth-grade test scores and misbehavior are controlled for. Yet this is not the case: once eighth-grade test scores and misbehavior are controlled for, the statistical and economic significance of these family characteristics is essentially zero.⁴⁴ Second, one would expect that youth with higher eighth-grade test scores (i.e., who have acquired more cognitive human capital by eighth grade) would cope better – in other words, would have higher tenth-grade test scores – even if they misbehaved in eighth grade. Column 6 investigates this issue by allowing the coefficients for eighth-grade test scores and misbehavior to vary independently for those students with above-average eighth-grade test scores. There is no evidence in the data supporting the conjecture that higher initial levels of cognitive human capital reduce the association between eighth-grade misbehavior and tenth-grade test scores. If anything, misbehavior in school most hurts those with above-average eighth-grade test scores. Hence, it seems that those youth who have accumulated more cognitive human capital by eighth grade pay the highest price, in terms of their tenth-grade test scores, for misbehavior. This result suggests that lower human capital levels are caused by school misbehavior that is continuing and not limited to the eighth grade. Thus, even misbehaving students with above-average eighth-grade test scores cannot catch up with their well-behaved peers and so will continue to accumulate human capital at lower rates.

5.3. Person-specific Type

The last explanation suggested for the empirical findings in this paper is the existence of a person-specific type that is the origin of misbehavior in eighth grade. Specifically, I suggest that eighth-grade misbehavior proxies for personality traits that are important in the labor market.

The empirical evidence suggests that employers value workers who display good work habits and that they are also willing to expend resources to avoid hiring workers who will display counterproductive work behaviors (Bishop, 1993; Bewley, 1999;

43. Note that if there were two groups of boys, the permanently-misbehaving and the misbehaving-in-eighth-grade-only, then if one-shot misbehavior had no effect on future outcomes then the results obtained previously would be biased downward.

44. Recall that there is a positive association between family income and educational attainment. Therefore, if resource-rich families help their boys catch up but disadvantaged families cannot (and hence their sons drop out), then we should expect to find no relationship between eighth-grade misbehavior and tenth-grade test scores in the selected sample used in Table 6.

Roberts et al., 2007; Autor and Scarborough, 2008).⁴⁵ Roberts et al. (2007) find that individuals who were less conscientious at 18 were more likely to engage in counterproductive work behavior, which includes being late to work, pretending to be sick, having conflicts with the boss, and doing things that could get them fired.⁴⁶ Therefore, it is possible that the personality trait proxied by eighth-grade misbehavior is conscientiousness.

Castillo et al. (2011) find that eighth-graders with higher discount rates have higher rates of disciplinary referrals. Consistent with the hypothesis that misbehavior is indicative of one's type or stable characteristics, the authors report that impatient eighth-graders are also more likely to have disciplinary referrals in ninth grade; this finding suggests that the discount rate is the link between eighth-grade misbehavior and outcomes. Specifically: since behaving well in school is costly and the benefits are not immediate, it could be that students with low discount rates do not find it worthwhile to acquire these costs and instead misbehave. It may likewise be that the same individuals do not find it worthwhile to bear the costs of behaving well in the workplace because they discount the future benefits to a large extent and hence display inappropriate behavior.

It is worth noting that these two interpretations are compatible. In particular, individuals who are more conscientious are more self-controlled and able to plan. Having a high discount rate means that one values the future and therefore is able to delay gratification (for high enough future rewards). Hence, it seems likely that individuals with high discount rates are more conscientious.⁴⁷

6. Conclusions

In this paper I exploit the NELS data set to investigate the associations between childhood behavior and cognitive human capital – as measured, respectively, by misbehavior in school and performance on achievement tests – with adult economic success. The measures of childhood misbehavior are based on detailed teacher

45. The value to the employer could stem from higher productivity related to good behavior, from higher morale in the workplace that in turn increases productivity as in Bewley (1999), or from reducing enforcement costs (as in Bowles et al., 2001b). This latter explanation is in agreement with the extended model, as those with low values of β (i.e., noncognitive skills) value the future benefits of keeping their job to a lesser degree and hence are more likely to shirk.

46. The authors also find that individuals rated low on agreeableness - a term that reflects individual differences in the tendency to be altruistic, trusting, modest, and warm (Roberts et al., 2004) – and in particular, those who are highly aggressive, are more likely to display counterproductive behaviors in the workplace.

47. Ameriks et al. (2007) use the gap between ideal and expected consumption as an economic measure of self-control and temptation avoidance and show that it is positively related to both wealth accumulation and conscientiousness.

evaluations of eighth-graders' behavior. That allows me to avoid problems associated with noncognitive skills measures that are self-reported, reflective, or that are concurrent with the outcomes of interest, which may have biased results in previous studies. The good behavior required in school seems, moreover, strongly similar to the good behavior valued in the workplace, and evidence in the literatures of economics and psychology suggest that misbehavior in school is related not only to low discount rates but also to low conscientiousness.

The empirical analysis shows the importance of good behavior in eighth grade for economic success beyond the impact of eighth-grade test scores. Eighth-grade misbehavior is associated with lower educational attainment even after controlling for eighth-grade test scores and family background characteristics. Misbehavior in eighth grade is negatively correlated with earnings. Controlling in addition for educational attainment, I find a direct relationship between eighth-grade misbehavior and earnings that accounts for nearly two thirds of the overall correlation. Furthermore, misbehavior in eighth grade is associated with earnings for all workers whereas performance on the eighth-grade achievement test is associated with earnings only for workers who obtain postsecondary degree. These findings are not driven by boys' tastes for future occupation or schooling. These associations are due in part to misbehaving boys accumulating less knowledge. However, it is probable that eighth-graders who misbehave continue to do so both in school and in the workplace and that this behavior explains most of the associations in the data.

The results indicate that, at least after eighth grade, the production function of cognitive human capital does not directly depend on the noncognitive skills examined in this paper. Instead, the dependence arises from high discounting of the future, which results in little effort investment in schooling and in misbehavior. The results also imply to an additive in effort production function and to an effort cost function that does not depend directly on any form of human capital. Of course, these results may not be applicable to other forms of noncognitive skills that may affect human capital accumulation and economic success through different mechanisms. Nor do they suggest that the production function of cognitive human capital is unchangeable at earlier ages.

The forms of misbehavior investigated in this paper may seem harmless, but the findings suggest that misbehavior in the classroom has substantial negative associations with boys' future earnings and educational attainment. Given the large magnitude of the associations found between misbehavior and outcomes and given that more than half of the boys in the NELS sample misbehave, it is tempting to suggest policies that target misbehavior in addition to, or even instead of, cognitive skills. The evidence in Segal (2008) indicates that, at least after eighth grade, misbehavior in school is quite persistent. It is possible that behavior in school at younger ages is more malleable; the evidence in Cunha and Heckman (2008) suggests

that in early childhood both cognitive and noncognitive skills are malleable and that noncognitive skills remain malleable until later ages. Without a better understanding of childhood misbehavior and the channels through which the associations between school misbehavior and outcomes arise, any policy recommendation would be premature. More data on childhood behavior and its relation to adult outcomes would shed light on some of these open questions.

References

- Ameriks, John, Andrew Caplin, John Leahy, and Tom Tyler (2007). “Measuring Self-Control Problems.” *American Economic Review*, 97, 966–972.
- Autor, David H. and David Scarborough (2008). “Does Job Testing Harm Minority Workers? Evidence from Retail Establishments.” *The Quarterly Journal of Economics*, 123, 219–277.
- Barbaranelli, Claudio, Gian Vittorio Caprara, Annarita Rabasca, and Concetta Pastorelli (2003). “A questionnaire for measuring the Big Five in late childhood.” *Personality and Individual Differences*, 34, 645–664.
- Barro, Robert J. and Rachel M. McCleary (2003). “Religion and Economic Growth across Countries.” *American Sociological Review*, 68, 760–781.
- Becker, Gary S. and Casey B. Mulligan (1997). “The Endogenous Determination of Time Preference.” *The Quarterly Journal of Economics*, 112, 729–758.
- Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (forthcoming). “Who is ‘Behavioral’? Cognitive Ability and Anomalous Preferences.” *Journal of the European Economic Association*.
- Bewley, Truman (1999). *Why Wages Don’t Fall During a Recession*. Harvard University Press.
- Bishop, John (1993). “Improving Job Matches in the U.S. Labor Market.” *Brookings Papers on Economic Activity. Microeconomics*, 1993, 335–400.
- Bowles, Samuel and Herbert Gintis (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. Basic Books.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne (2001a). “The Determinants of Earnings: A Behavioral Approach.” *Journal of Economic Literature*, 39, 1137–1176.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne (2001b). “Incentive-Enhancing Preferences: Personality, Behavior, and Earnings.” *American Economic Review*, 91, 155–158.
- Cameron, Stephen V. and James Heckman (2001). “The Dynamics of Educational Attainment for Black, Hispanic, and White Males.” *Journal of Political Economy*, 109, 455–499.

- Carneiro, Pedro and James Heckman (2003). "Human Capital Policy." In *Inequality in America: What Role for Human Capital Policy?*, edited by A. Krueger and J. Heckman, pp. 77–240. MIT Press, Massachusetts.
- Castillo, Marco, Paul Ferraro, Jeff Jordan, and Ragan Petrie (2011). "The Today and Tomorrow of Kids." *Journal of Public Economics*, 95, 1377–1385.
- Cawley, John, James Heckman, and Edward Vytlacil (2001). "Three observations on wages and measured cognitive ability." *Labour Economics*, 8, 419–442.
- Cunha, Flavio and James Heckman (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources*, 43, 738–814.
- Dee, Thomas S. (2007). "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources*, 42, 528–554.
- DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47, 315–372.
- Dodge, K.A. (2006). "Translational Science in Action: Hostile Attributional Style and the Development of Aggressive Behavior Problems." *Development and Psychopathology*, 18, 791–814.
- Doepke, Matthias and Fabrizio Zilibotti (2008). "Occupational Choice and the Spirit of Capitalism." *The Quarterly Journal of Economics*, 123, 747–793.
- Dohmen, Thomas J, Armin Falk, David Huffman, and Uwe Sunde (2011). "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100, 1238–1260.
- Dreeben, Robert (1976). *On What Is Learned in School*. Addison-Wesley.
- Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer (1995). "Do Teachers' Race, Gender, and Ethnicity Matter?: Evidence from NELS88." *Industrial and Labor Relations Review*, 48, 547–556.
- Grogger, Jeffrey and Derek Neal (2000). "Further Evidence on the Effects of Catholic Secondary Schooling." *Brookings-Wharton Papers on Urban Affairs*, 1, 151–201.
- Heckman, James (2000). "Policies to foster human capital." *Research in Economics*, 54, 3–56.
- Heckman, James and Yona Rubinstein (2001). "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review*, 91, 145–149.
- Heckman, James, Jora Stixrud, and Sergio Urzua (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24, 411–482.
- Ichino, Andrea and Enrico Moretti (2009). "Biological Gender Differences, Absenteeism, and the Earnings Gap." *American Economic Journal: Applied Economics*, 1, 183–218.

- Jacob, Brian A. (2002). "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*, 21, 589–598.
- Kuhn, Peter and Catherine Weinberger (2005). "Leadership Skills and Wages." *Journal of Labor Economics*, 23, 395–434.
- Lavy, Victor (2008). "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics*, 92, 2083–2105.
- Lazear, Edward P. (2009). "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Public Economics*, 117, 914–940.
- Lindqvist, Erik and Roine Westman (2011). "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment." *American Economic Journal: Applied Economics*, 3, 101–128.
- Mulligan, Casey B. and Yona Rubinstein (2008). "Selection, Investment, and Women's Relative Wages Over Time." *Quarterly Journal of Economics*, 123, 1061–1110.
- Ones, Deniz S., Shtephan Dilchert, Chockalingam Viswesvaran, and Timothy A. Judge (2007). "In support of personality assessment in organizational settings." *Personnel Psychology*, 60, 995–1027.
- Parsons, Talcott (1959). "The School Class as a Social System." *Harvard Educational Review*, 29, 297–318.
- Persico, Nicola, Andrew Postlewaite, and Dan Silverman (2004). "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height." *Journal of Political Economy*, 112, 1019–1053.
- Roberts, Brent W., Peter D. Harms, Avshalom Caspi, and Terri E. Moffitt (2007). "Predicting the Counterproductive Employee in a Child-to-Adult Prospective Study." *Journal of Applied Psychology*, 92, 1427–1436.
- Roberts, Brent W., Richard W. Robins, Kali H. Trzesniewski, and Avshalom Caspi (2004). "Personality Trait Development in Adulthood." In *Handbook of the Life Course*, edited by Michael J. Shanahan Jeylan T. Mortimer, pp. 579–598. Springer, New York.
- Segal, Carmit (2005). "Esseys on Human Capital." Ph.D. dissertation, Stanford University.
- Segal, Carmit (2008). "Classroom Behavior." *Journal of Human Resources*, 43, 783–814.
- Segal, Carmit (forthcoming). "Working When No One Is Watching: Motivation, Test Scores, and Economic Success." *Management Science*.

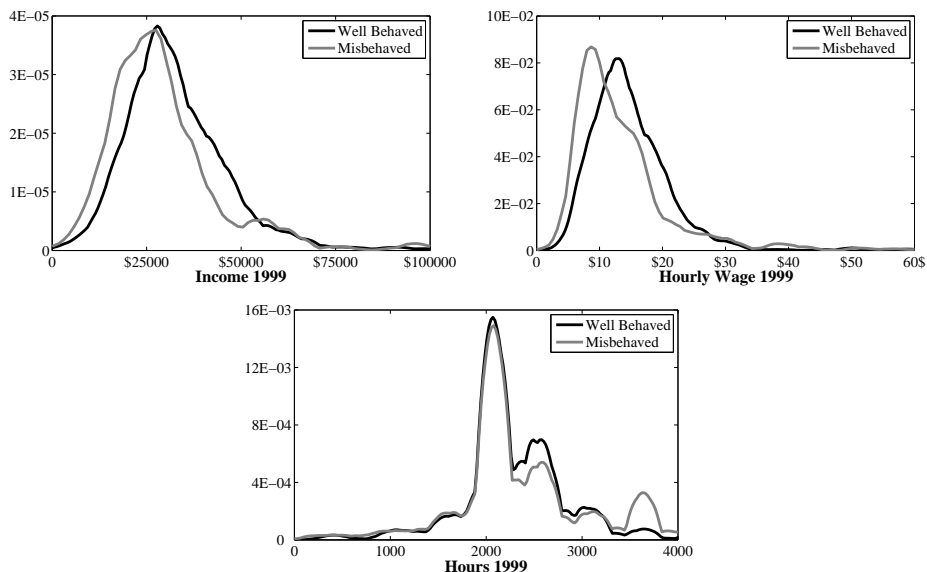


Figure 1. Kernel densities for the two types of behavior (where “Well Behaved” is a dummy set equal to 1 only if both teachers reported good behavior in all five categories; see text for details).

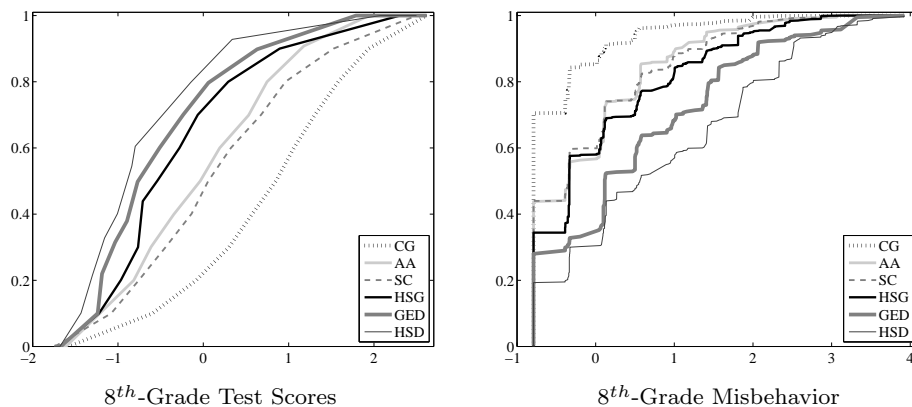


Figure 2. The weighted cumulative distribution of eighth-grade test scores and misbehavior by educational attainment as of 2000 (where the misbehavior variable used is the composite one; see text for details).

TABLE 1. Means of key outcome variables by eighth-grade misbehavior.

Variable	Definition	Mean	Well Behaved ^a	Misbehaved ^b	Observations
Test Scores					
Eighth-Grade	Standardized 8 th -grade math/reading test scores (see Section 3)	0 (1)	0.39 (0.98)	-0.32 (0.90)	9,335
Tenth-Grade	Standardized 10 th -grade math/reading test scores (see Section 3)	0.1 (0.99)	0.39 (0.93)	-0.24 (0.95)	3,440
Educational Attainment by 2000 (for males with nonmissing data on all family characteristics)					
HSD	A dummy equal to 1 if the person is a high school dropout	0.04	0.03	0.10	3,852
GED	A dummy equal to 1 if a GED was obtained	0.04	0.04	0.08	3,852
HSG	A dummy equal to 1 if the person is at most high school graduate	0.21	0.17	0.32	3,852
SC	A dummy equal to 1 if the person was enrolled in a postsecondary institution, but has not obtained any degree	0.30	0.26	0.28	3,852
AA ^c	A dummy equal to 1 if an Associate of Arts degree was obtained	0.07	0.06	0.07	3,852
CG	A dummy equal to 1 if at least a bachelor degree was obtained	0.34	0.43	0.16	3,852
Earnings related variables in 1999 (for full-time male workers who were not self-employed and not in school)					
Income ^d	The income from employment	\$31,635 (\$17,424)	\$33,406 (\$14,589)	\$30,249 (\$19,248)	2,240
Hours ^d	The number of hours worked, which equals the average hours per week multiplied by the number of weeks worked.	2,361 (592)	2,317 (506)	2,395 (649)	2,240
Hourly Wage ^d	The average hourly wage, which equals Income divided by Hours	\$13.95 (\$8.47)	\$14.84 (\$7.25)	\$13.25 (\$9.25)	2,240

Notes: Standard errors in parentheses. Reported values weighted by the NELS sampling weights. The samples include only men with eighth-grade test scores and two teacher evaluations, and they correspond to the samples used in the analysis. Specifically, the tenth-grade test scores correspond to the Table 6 sample, educational attainment corresponds to the Table 2 sample, and earnings-related variables correspond to the sample used in Tables 3 and 4.

a. Both eighth-grade teachers reported good behavior in all five categories.

b. At least one eighth-grade teacher reported misbehavior in at least one of the five categories.

c. To distinguish individuals who failed to get an academic degree from those who failed to get any kind of license/certificate, "postsecondary institution" is defined as a two-year public, four-year public, or four-year private (not-for-profit) postsecondary institution.

d. Two observations—an hourly wage of \$1,143 and one of 3 cents—were deemed to be outliers and hence dropped from the sample.

TABLE 2. Relationship between eighth-grade misbehavior and the probability of obtaining a postsecondary degree by 2000: marginal effects after probit, evaluated at the sample means.

	[1]	[2]	[3]
Test Scores	0.109 [0.016]***	0.107 [0.016]***	0.096 [0.018]***
Misbehavior	-0.086 [0.021]***		-0.091 [0.019]***
Tardy ^a		0.076 [0.076]	
Absent ^a		-0.099 [0.039]**	
Disruptive ^a		-0.049 [0.037]	
Does Not Complete Homework ^a		-0.126 [0.038]***	
Inattentive ^a		-0.02 [0.040]	
Test Scores × Misbehavior ^b			-0.038 [0.018]**
Black ^a	0.032 [0.070]	0.039 [0.070]	0.03 [0.068]
Hispanic ^a	0.045 [0.048]	0.035 [0.048]	0.044 [0.048]
Father High School Dropout ^a	-0.105 [0.047]**	-0.091 [0.046]**	-0.102 [0.045]**
Father College Graduate or More ^a	0.09 [0.038]**	0.086 [0.038]**	0.088 [0.037]**
Mother High School Dropout ^a	-0.071 [0.041]*	-0.078 [0.042]*	-0.069 [0.040]*
Mother College Graduate or More ^a	0.142 [0.033]***	0.147 [0.032]***	0.142 [0.033]***
Family Income (in \$10,000)	0.021 [0.005]***	0.021 [0.005]***	0.02 [0.005]***
Does Not Live with Both Parents ^a	-0.143 [0.030]***	-0.135 [0.030]***	-0.137 [0.029]***
Number of Siblings	-0.017 [0.012]	-0.017 [0.011]	-0.018 [0.011]
Joint Test: Behavior Prob. > χ^2		41.3 <0.001	
Observations	3,852	3,852	3,852
Predicted Prob. at the Sample Means	0.311	0.031	0.302

Notes: Robust standard errors (clustered by eighth-grade school) in brackets. The sample is restricted to include only those students with data on test scores, two teacher evaluations, and data on all the family background characteristics. Misbehavior measures and Test Scores are for the eighth grade.

a. The marginal effect is for a discrete change of a dummy variable from 0 to 1

b. Marginal effect for this interaction term is evaluated at the respective mean of each variable.

Significant at 10%; **significant at 5%; ***significant at 1%.

TABLE 3. Relationship between eighth-grade misbehavior and earnings in 1999 (full-time workers).

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Test Scores		0.076	0.055	0.053	0.013	0.012			
		[0.018]***	[0.018]***	[0.018]***	[0.020]	[0.021]			
Misbehavior			-0.062		-0.041			-0.044	
			[0.016]***		[0.018]**			[0.017]**	
Test Scores – Postsecondary Degree							0.078	0.072	0.068
							[0.027]***	[0.026]***	[0.026]***
Test Scores – No Postsecondary Degree							-0.008	-0.020	-0.020
							[0.026]	[0.026]	[0.026]
Tardy				-0.110		-0.132			
				[0.067]*		[0.068]*			
Absent				-0.049		-0.014			
				[0.058]		[0.059]			
Disruptive				0.075		0.075			
				[0.039]*		[0.039]*			
Does Not Complete Homework				-0.071		-0.037			
				[0.053]		[0.053]			
Inattentive				-0.075		-0.061			
				[0.052]		[0.052]			
Black		-0.130	-0.077	-0.073	-0.086	-0.079	-0.088	-0.094	-0.087
		[0.062]**	[0.065]	[0.063]	[0.061]	[0.061]	[0.063]	[0.059]	[0.060]
Hispanic		-0.132	-0.088	-0.072	-0.070	-0.067	-0.084	-0.073	-0.070
		[0.072]*	[0.070]	[0.069]	[0.067]	[0.069]	[0.069]	[0.069]	[0.071]
Completed Years of Schooling					0.048	0.048	0.048	0.043	0.043
					[0.011]***	[0.011]***	[0.011]***	[0.011]***	[0.011]***
Constant	10.281	10.271	10.266	10.318	9.641	9.667	9.635	9.694	9.723
	[0.018]***	[0.018]***	[0.018]***	[0.019]***	[0.149]***	[0.155]***	[0.145]***	[0.145]***	[0.152]***
Joint Test: Behavior				4.89		2.73			2.88
Prob. > F				<0.001		0.019			0.014
Observations	2,240	2,240	2,240	2,240	2,240	2,240	2,240	2,240	2,240
R ²	0.005	0.03	0.05	0.05	0.08	0.09	0.08	0.09	0.09

Notes: Robust standard errors (clustered by eighth-grade school) in brackets. The sample is restricted to include only those students (with data on test scores and two teacher evaluations) who reported working at least one hour per week for at least one week in 1999, were not in school, and were not self-employed. Misbehavior measures and Test Scores are for the eighth grade.

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 4. Eighth-grade misbehavior and earnings in 1999 (full-time workers): testing for production function functional form

	Dependent Variable: ln(Earnings(1999))						
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Test Scores – Postsecondary Degree	0.086 [0.033]***	0.074 [0.028]***	0.086 [0.033]***	0.086 [0.034]**	0.074 [0.029]***	0.086 [0.034]**	0.071 [0.025]***
Test Scores – No Postsecondary Degree	-0.021 [0.027]	-0.02 [0.027]	-0.021 [0.027]	-0.021 [0.026]	-0.02 [0.026]	-0.021 [0.026]	0.02 [0.027]
Misbehavior – Postsecondary Degree	-0.043 [0.036]	-0.045 [0.039]	-0.043 [0.036]	-0.043 [0.036]	-0.043 [0.036]	-0.043 [0.039]	-0.045 [0.039]
Misbehavior – No Postsecondary Degree	-0.044 [0.022]**	-0.042 [0.021]**	-0.044 [0.018]**	-0.044 [0.019]**	-0.042 [0.019]**	-0.044 [0.017]**	-0.044 [0.018]**
Misbehavior							
Test Scores×Misbehavior – Postsecondary Degree	0.031 [0.044]		0.031 [0.044]	0.031 [0.044]		0.031 [0.044]	
Test Scores×Misbehavior – No Postsecondary Degree	-0.002 [0.024]		-0.002 [0.024]	-0.001 [0.023]			
Test Scores×Misbehavior		0.005 [0.021]			0.005 [0.021]		
Black	-0.096 [0.059]	-0.094 [0.059]	-0.096 [0.059]	-0.096 [0.059]	-0.094 [0.059]	-0.096 [0.059]	-0.094 [0.059]
Hispanic	-0.074 [0.069]	-0.073 [0.069]	-0.073 [0.069]	-0.074 [0.069]	-0.073 [0.069]	-0.073 [0.069]	-0.073 [0.069]
Completed Years of Schooling	0.044 [0.012]***	0.043 [0.011]***	0.044 [0.012]***	0.044 [0.011]***	0.043 [0.011]***	0.044 [0.011]***	0.043 [0.012]***
Constant	9.681 [0.156]***	9.695 [0.152]***	9.681 [0.154]***	9.681 [0.148]***	9.694 [0.144]***	9.682 [0.147]***	9.695 [0.152]***
Tests for the Equality of Coefficients across Education Levels (Prob. > F)							
Test Scores Coefficients	0.009	0.009	0.009	0.009	0.009	0.009	0.008
Misbehavior Coefficients	0.978	0.941	0.99				0.975
Test Scores×Misbehavior Coefficients	0.517			0.521			
Observations	2,240	2,240	2,240	2,240	2,240	2,240	2,240
R ²	0.09	0.09	0.09	0.09	0.09	0.09	0.09

See notes to Table 3

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 5. Eighth-grade misbehavior and earnings (full-time workers): expected occupation at age 30 and expected education.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Dependent Variable: ln(Earnings(1999))							
Test Scores – Postsecondary Degree	0.060 [0.029]**	0.059 [0.029]**	0.065 [0.029]**	0.064 [0.029]**	0.071 [0.026]***	0.068 [0.026]***	0.057 [0.025]**	0.055 [0.025]**
Test Scores – No Postsecondary Degree	-0.034 [0.030]	-0.029 [0.030]	-0.036 [0.028]	-0.030 [0.028]	-0.021 [0.026]	-0.021 [0.026]	-0.037 [0.027]	-0.036 [0.027]
Misbehavior	-0.045 [0.023]*		-0.051 [0.021]**		-0.045 [0.017]**		-0.041 [0.017]**	
Tardy		-0.203 [0.080]**		-0.182 [0.069]**		-0.126 [0.066]*		-0.121 [0.063]*
Absent		-0.052 [0.073]		-0.067 [0.058]		-0.023 [0.059]		-0.021 [0.056]
Disruptive		0.041 [0.051]		0.033 [0.049]		0.074 [0.039]*		0.070 [0.039]*
Does Not Complete Homework		0.023 [0.059]		0.021 [0.057]		-0.041 [0.052]		-0.035 [0.051]
Inattentive		-0.034 [0.063]		-0.042 [0.062]		-0.063 [0.052]		-0.057 [0.051]
Black	-0.096 [0.069]	-0.113 [0.070]	-0.096 [0.070]	-0.114 [0.070]	-0.096 [0.059]	-0.088 [0.060]	-0.125 [0.061]**	-0.117 [0.061]*
Hispanic	-0.122 [0.088]	-0.117 [0.087]	-0.135 [0.078]*	-0.129 [0.076]*	-0.074 [0.069]	-0.070 [0.071]	-0.068 [0.067]	-0.065 [0.069]
Completed Years of Schooling	0.040 [0.014]***	0.044 [0.014]***	0.036 [0.014]***	0.039 [0.013]***	0.043 [0.011]***	0.043 [0.011]***	0.036 [0.011]***	0.037 [0.011]***
Expected Years of Schooling							0.025 [0.009]***	0.024 [0.009]***
Constant	9.759 [0.183]***	9.726 [0.191]***	9.771 [0.177]***	9.763 [0.178]***	9.696 [0.145]***	9.726 [0.152]***	9.408 [0.154]***	9.443 [0.167]***
Joint Test: Behavior Prob. > F		2.24		2.28		2.95		2.68
Expected Occupation at Age 30 Dummies	No	No	Yes	Yes	No	No	No	No
Observations	1,597	1,597	1,597	1,597	2,234	2,234	2,234	2,234
R ²	0.08	0.09	0.11	0.11	0.09	0.10	0.09	0.10

Note: See notes to Table 3.

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 6. Relation between tenth-grade test scores and eighth-grade misbehavior.

	Dependent Variable: Tenth-Grade Test Scores					
	[1]	[2]	[3]	[4]	[5]	[6]
Eighth-Grade Test Scores				0.820 [0.012]***	0.818 [0.012]***	0.913 [0.028]***
8 th Grade Test Scores: Above-Average 8 th Grade Test Scores						-0.160 [0.041]***
Eighth-Grade Misbehavior		-0.312 [0.027]***		-0.083 [0.015]***		-0.059 [0.018]***
8 th Grade Misbehavior: Above-Average 8 th Grade Test Scores Tardy			0.053 [0.096]		-0.000 [0.047]	-0.057 [0.028]**
Absent			-0.073 [0.067]		-0.007 [0.034]	
Disruptive			-0.174 [0.051]***		-0.047 [0.027]*	
Does Not Complete Homework			-0.284 [0.051]***		-0.080 [0.028]***	
Inattentive			-0.267 [0.055]***		-0.068 [0.027]**	
Black	-0.442 [0.082]***	-0.391 [0.080]***	-0.385 [0.080]***	-0.082 [0.041]**	-0.080 [0.041]**	-0.074 [0.041]*
Hispanic	-0.249 [0.092]***	-0.225 [0.087]***	-0.206 [0.087]**	-0.106 [0.041]***	-0.100 [0.041]**	-0.104 [0.040]**
Father High School Dropout	-0.121 [0.065]*	-0.094 [0.065]	-0.087 [0.064]	-0.016 [0.032]	-0.015 [0.033]	-0.015 [0.032]
Father College Grad or More	0.321 [0.051]***	0.276 [0.049]***	0.267 [0.049]***	0.029 [0.027]	0.027 [0.027]	0.025 [0.027]
Mother High School Dropout	-0.179 [0.067]***	-0.176 [0.066]***	-0.181 [0.065]***	-0.022 [0.034]	-0.024 [0.034]	-0.024 [0.034]
Mother College Grad or More	0.163 [0.050]***	0.147 [0.047]***	0.148 [0.048]***	0.018 [0.027]	0.018 [0.027]	0.026 [0.027]
Family Income in \$10,000	0.013 [0.007]*	0.012 [0.007]*	0.011 [0.007]	0.002 [0.003]	0.001 [0.003]	0.002 [0.003]
Does Not Live with Both Parents	-0.163 [0.047]***	-0.109 [0.047]**	-0.111 [0.047]**	-0.081 [0.023]***	-0.082 [0.023]***	-0.078 [0.022]***
Number of Siblings	-0.030 [0.014]**	-0.025 [0.014]*	-0.025 [0.014]*	-0.007 [0.007]	-0.008 [0.007]	-0.008 [0.007]
Constant	0.105 [0.049]**	0.018 [0.049]	0.299 [0.049]***	-0.070 [0.024]***	0.006 [0.026]	-0.015 [0.030]
Joint Test: Behavior Prob. > F			36.35 <0.001		7.79 <0.001	
School Dummies ^a	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,440	3,340
R ²	0.43	0.47	0.48	0.84	0.84	0.85

Notes: Robust standard errors in brackets. The sample is restricted to include only those students with data on test scores, two teacher evaluations in both grades, and data on all the family background characteristics. The sample is further restricted to include only those eighth-grade schools containing at least two students with valid data. Misbehavior measures are for eighth grade.

a. Jointly significant in all specifications.

* significant at 10%; ** significant at 5%; *** significant at 1%