DISCUSSION PAPER SERIES

IZA

IZA DP No. 8027

Education, Health and Wages

James J. Heckman
John Eric Humphries
Gregory Veramendi
Sergio Urzúa

March 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Education, Health and Wages

James J. Heckman
University of Chicago,
American Bar Foundation and IZA

John Eric Humphries
University of Chicago

Gregory Veramendi
Arizona State University

Sergio Urzúa
University of Maryland
NBER and IZA

Discussion Paper No. 8027
March 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8027
March 2014

ABSTRACT

Education, Health and Wages*

This paper develops and estimates a model with multiple schooling choices that identifies the causal effect of different levels of schooling on health, health-related behaviors, and labor market outcomes. We develop an approach that is a halfway house between a reduced form treatment effect model and a fully formulated dynamic discrete choice model. It is computationally tractable and identifies the causal effects of educational choices at different margins. We estimate distributions of responses to education and find evidence for substantial heterogeneity in unobserved variables on which agents make choices. The estimated treatment effects of education are decomposed into the direct benefits of attaining a given level of schooling and indirect benefits from the option to continue on to further schooling. Continuation values are an important component of our estimated treatment effects. While the estimated causal effects of education are substantial for most outcomes, we also estimate a quantitatively important effect of unobservables on outcomes. Both cognitive and socioemotional factors contribute to shaping educational choices and labor market and health outcomes. We improve on LATE by identifying the groups affected by variations in the instruments. We find benefits of cognition on most outcomes apart from its effect on schooling attainment. The benefits of socioemotional skills on outcomes beyond their effects on schooling attainment are less precisely estimated.

JEL Classification: C32, C38, I12, I14, I21

Keywords: education, early endowments, factor models, health, treatment effects

Corresponding author:

James Heckman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL
USA
E-mail: jjh@uchicago.edu

* We thank the editor, Rob Shimer, two anonymous referees, and Chris Taber for very helpful comments on this draft. This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, the Buffett Early Childhood Fund, NICHD 5R37HD065072, 1R01HD54702, an anonymous funder, the support of a European Research Council grant hosted by University College Dublin, DEVHEALTH 269874, and a grant from the Institute for New Economic Thinking (INET) to the Human Capital and Economic Opportunity Global Working Group (HCEO)-an initiative of the Becker Friedman Institute for Research in Economics (BFI). Humphries acknowledges the support of a National Science Foundation Graduate Research Fellowship. The views expressed in this paper are those of the authors and not necessarily those of the funders or persons named here. The Web Appendix for this paper is http://heckman.uchicago.edu/effect_ed_choice_labor.

1 Introduction

Persons with more schooling have better health, healthier behaviors, and higher wages. The causal basis for these relationships is the topic of a large literature.[1]

Figure 1 illustrates these familiar empirical regularities for the four outcomes analyzed in this paper: wages, health, self-esteem, and smoking.[2] The black bar in each panel shows the unadjusted mean difference in outcomes for persons at the indicated levels of educational attainment compared to those for high school dropouts. An obvious explanation for such relationships is ability bias. However, as shown by the grey bars in Figure 1, adjusting for family background and adolescent measures of ability attenuates, but does not eliminate, the estimated effects of education.[3]

It is easy to fault the simple adjustments used in Figure 1. The measures used to make the adjustments may be incomplete or imperfect. Such concerns have given rise to a search for instrumental variables (exclusion restrictions including randomization and regression discontinuity methods) to secure causal estimates of the effect of education.

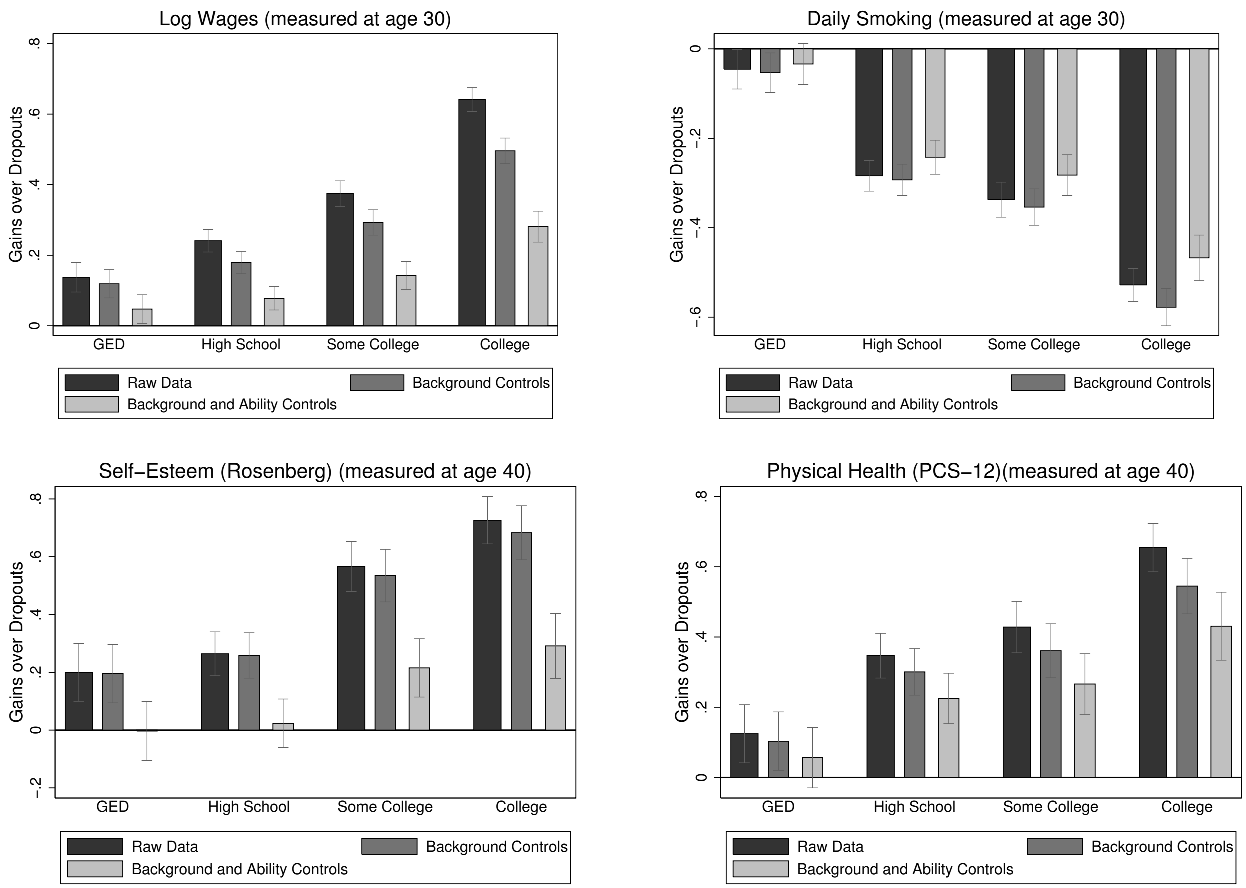
The IV literature has itself been faulted. The available instruments for schooling are often weak (see Bound, Jaeger, and Baker, 1995). When agents select into schooling on the basis of their idiosyncratic benefits from it, instruments identify "causal effects" for unidentified groups of persons (see, e.g., Heckman, 1997; Heckman and Vytlacil, 1999, 2005) that often differ from average treatment effects, treatment on the treated, or policy-relevant treatment effects (see Carneiro, Heckman, and Vytlacil, 2011). Regression discontinuity methods identify responses for individuals at the points of discontinuity of the instruments which may or may not be the individuals toward which policies are best targeted.

[1] A positive association between education and labor market outcomes has long been noted in the literature (Mincer, 1958; Becker, 1964; Mincer, 1974). For surveys, see Card (1999) and Heckman, Lochner, and Todd (2006) and the references they cite. The positive correlation between schooling and health is also a well-established finding (Grossman, 1972, 2000, 2006). See also Adams (2002); Arendt (2005); Lleras-Muney (2005); Silles (2009); Spasojevic (2003); Arkes (2003); Auld and Sidhu (2005); Grossman (2008); Grossman and Kaestner (1997); Cutler and Lleras-Muney (2010); Conti, Heckman, and Urzua (2010), and the literature review in Section B of the Web Appendix.

[2] The Web Appendix reports results for other outcomes.

[3] See, e.g., the papers cited in Card (1999).

Figure 1: The Observed Benefits from Education after Controlling for Background and Ability



Notes: The bars represent the coefficients from a regression of the designated outcome on dummies for educational attainment, where the omitted category is high school dropout. Regressions are run adding successive controls for background and proxies for ability. Background controls include race, region of residence in 1979, urban status in 1979, broken home status, number of siblings, mother's education, father's education, and family income in 1979. Proxies for ability are average score on the ASVAB tests and ninth grade GPA in core subjects (language, math, science, and social science). "Some College" includes anyone who enrolled in college, but did not receive a four-year college degree.
Source: NLSY79 data.

Most of the treatment effect literature focuses on identifying the effects of choosing between two levels of final schooling attainment or else assumes that schooling is captured by "years of schooling"[4] and estimates versions of the Mincer (1974) model.[5] With a few notable exceptions,[6] little work in the treatment effect literature considers models with multiple discrete schooling levels or dynamic models of schooling attainment. Moreover, identifying treatment effects at multiple margins of choice requires choice-specific instruments that are often not available.

A growing literature formulates and estimates dynamic discrete choice models of schooling that account for both the nonlinearity of the effects of schooling and the information available to agents when they make their schooling choices.[7] With these models, it is possible to identify the margins of choice which different instruments identify and the populations affected by the various instruments.[8]

However, many question the robustness of estimates from such models because of the often strong assumptions made about schooling choice models, the information sets that agents are assumed to act on, the apparent arbitrariness in the choices of functional forms in the estimation equations, and the invocation of assumptions about the support of the instruments (see, e.g., Imbens, 2010). Many scholars report difficulties in identifying crucial cost parameters.[9] Nonpecuniary or "psychic" costs play a dominant role in many structural models of schooling in which agents are assumed to make choices to maximize expected future net income. See, e.g., Cunha, Heckman, and Navarro (2005); Eisenhauer, Heckman, and Mosso (2013); Abbott, Gallipoli, Meghir, and Violante (2013). Unexplained "psychic costs" or tastes for schooling substantially outweigh financial costs in accounting for schooling choices. This casts some doubt on

[4]See Card (1999) for a survey.

[5]See Heckman, Lochner, and Todd (2006) for a discussion of the empirical evidence against the Mincer model. There is abundant evidence of "sheepskin" effects, i.e. nonlinearities associated with completion of college. Those authors also show that the original Mincer model ignores continuation values to education which we show to be an important component of the "true" effect of schooling.

[6]Angrist and Imbens (1995) and Heckman, Urzua, and Vytlacil (2006). The latter paper points out some difficulties in the economic interpretation of the decompositions reported in the former paper.

[7]Keane and Wolpin (1997); Keane, Todd, and Wolpin (2011).

[8]This approach produces conceptually clean models. See Heckman and Urzua (2010).

[9]See, e.g., Eisenhauer, Heckman, and Mosso (2013) and Eisenhauer, Heckman, and Vytlacil (2014).

the specifications of decision rules used in the current structural literature.

As a result of the criticism directed against the various approaches, the literature on the "causal effects" of schooling is divided into camps organized around favored methodologies, as well as beliefs about the questions they think can be "credibly" answered by the data. This paper implements an approach that is a halfway house between the IV literature that reports "effects" at unspecified margins and the fully structural dynamic discrete choice literature. Our approach draws on identifying strategies from the matching, IV, and control function literatures.[10] We identify the causal effects of schooling at different stages of the life cycle based in part on a rich set of covariates to control for selection bias. We also use exclusion restrictions to identify our model as in the IV and control function literatures. Like the structural econometrics literature, we estimate causal effects at clearly identified margins of choice for populations affected by policies. Unlike the structural literature, we are agnostic about the specific model of choice used by agents. We approximate the dynamic choice model following suggestions of Heckman (1981), Eckstein and Wolpin (1989), Cameron and Heckman (2001), and Geweke and Keane (2001).

We build on the sequential discrete choice model of Cameron and Heckman (2001) by adding schooling-specific outcome equations and by adding interpretable measurements to proxy the cognitive and socioemotional variables found to be important predictors of schooling, the returns to schooling, and the psychic costs of schooling.[11] We explore the dimension of the space of unobservables required to control for selection and to fit the data. Instead of trying to purge the effects of multiple abilities on outcomes to isolate causal effects of schooling, we estimate the effects of these abilities in shaping schooling and in mediating the effects of schooling on outcomes. We use numerous proxies of both cognitive and socioemotional abilities in an attempt to account for ability bias at multiple margins of choice. We find that at least two dimensions of heterogeneity are required to produce an adequate empirical model of schooling and its effects on

[10] See Heckman (2008) for a review of these alternative approaches.

[11] See the evidence in Borghans, Duckworth, Heckman, and ter Weel (2008) and Almlund, Duckworth, Heckman, and Kautz (2011).

outcomes. We also find that after accounting for these abilities, there is little additional role for unmeasured abilities in shaping the dependence between schooling decisions and outcomes. We account for measurement error in measuring abilities and show that doing so has important consequences.[12] We capture heterogeneity in the response to treatment on which individuals sort into schooling that is a hallmark of the recent IV literature. We estimate the empirical consequences of sorting on multiple components of ability.

In our model, as in standard dynamic discrete choice models, educational choices at one stage open up educational options at later stages. The expected consequences of future choices and their costs are *implicitly* valued by individuals when deciding whether or not to continue their schooling. Our empirical strategy allows for these *ex ante* valuations but does not explicitly estimate them.[13] We decompose the *ex post* treatment effects of educational choices into the direct benefits of the choice and the continuation values arising from access to additional education beyond the current choice. Thus we estimate *ex post* returns to schooling both as the direct causal benefit comparing two final schooling levels—the traditional focus in the human capital literature (see, e.g., Becker, 1964)—and as returns through continuation values created by the options opened up by schooling (Weisbrod, 1962; Comay, Melnik, and Pollatschek, 1973; Altonji, 1993; Cameron and Heckman, 1993; and Heckman, Lochner, and Todd, 2006).

Our paper also contributes to an emerging literature on the importance of both cognitive and socioemotional skills in shaping life outcomes (see Borghans, Duckworth, Heckman, and ter Weel, 2008; Heckman, Stixrud, and Urzua, 2006; Almlund, Duckworth, Heckman, and Kautz, 2011). The traditional literature on the benefits of education focuses on the effects of cognitive ability. We confirm the findings in the recent literature that both cognitive and socioemotional skills are important predictors of educational attainment. Fixing schooling levels, the effects of cognition on outcomes are still substantial. The estimates of the within-schooling effects of socioemotional skills on outcomes are less precisely estimated.

[12]See Section F of the Web Appendix for evidence on this issue.

[13]See, e.g., Eisenhauer, Heckman, and Mosso (2013), where this is done.

We find that (a) there is substantial sorting into schooling both on cognitive and socioemotional measures; (b) there are causal effects of education on smoking, physical health, and wages at all levels of schooling; (c) for most outcomes, only high-ability people benefit from graduating from college. An exception to this rule is that low-ability people are the only ability group to benefit in terms of self-esteem; (d) continuation values are an important component of the causal effects of schooling; and (e) measurement error is empirically important, and ignoring it affects our estimates. We also contribute to the literature on the non-market benefits of education by studying the causal effects of education on health, healthy behaviors, and mental health (see, e.g., Lochner, 2011, Oreopoulos and Salvanes, 2011, and Cawley and Ruhm, 2012).

1.1 The Benefits and Limitations of Our Approach

Like the treatment effect literature, our approach enables us to identify the *gross benefits* of education. Unlike what is obtained from that approach, we can identify average benefits for persons at multiple levels of schooling in terms of observable and unobservable characteristics. The approach adopted in this paper complements the analyses of Heckman and Vytlacil (1999, 2005, 2007a,b); Heckman, Urzua, and Vytlacil (2006), and Carneiro, Heckman, and Vytlacil (2011) by providing a flexible parametric alternative to their data-demanding semi-parametric analyses. Our approach is useful in estimating parameters for samples of moderate size, such as the NLSY79 data analyzed in this paper.

Because we are agnostic about the decision model used by our agents, like the rest of the treatment effect literature, we do not identify costs.[14] We do not impose particular models of expectations such as rational expectations. Fully structural models do so, but typically impose greater parametric structure. Accordingly, we cannot estimate *ex ante*

[14]We can use auxiliary data to identify components of costs such as tuition. However, we cannot identify the full components of cost, including psychic costs, which have been estimated to be very important. See Eisenhauer, Heckman, and Mosso (2013), Eisenhauer, Heckman, and Vytlacil (2014), and Abbott, Gallipoli, Meghir, and Violante (2013) where explicit structural approaches are developed and applied to generate treatment effects and the costs and benefits of treatments. Structural models like those of Eisenhauer, Heckman, and Mosso (2013) identify both *ex ante* and *ex post* rates of return.

benefits and costs, nor can we estimate net rates of return. This limitation precludes a full cost-benefit analysis. With our approach we can only identify a portion of the ingredients required to evaluate social programs—the *ex post* gross benefit portion emphasized in the literature on treatment effects although monetary components of cost may be available from auxiliary sources. Our approach represents a computationally tractable compromise between the conventional literature on treatment effects and a fully structural approach that allows us to explore economically relevant margins of choice without imposing strong assumptions about agent decision making.

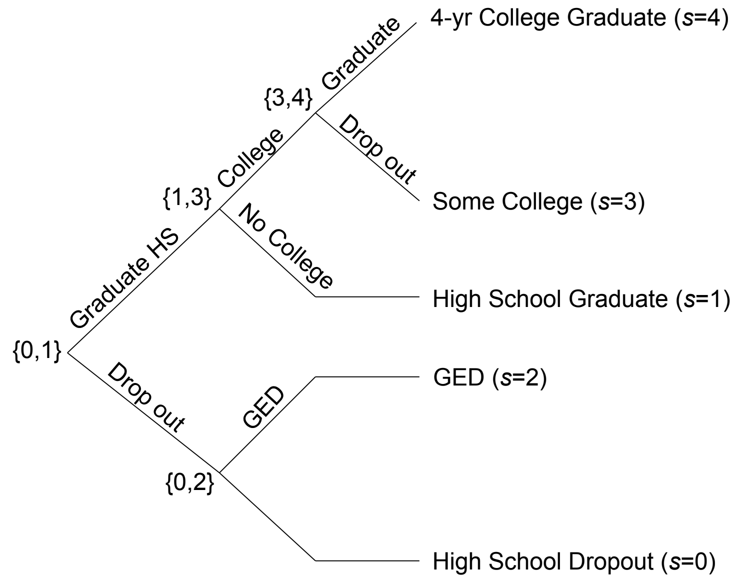
2 The Model

We estimate a sequential model of schooling with the transitions at the nodes shown in Figure 2, where $\mathcal{J}$ is a set of possible schooling states, $\mathcal{C}_{j,j'}$ is the available choice set for a person at j choosing between remaining at j or transiting to j', where $j, j' \in \mathcal{J}$. $\mathcal{J}$ is not necessarily ordered. $D_{j,j'} = 1$ if a person at j chooses $j' \in \mathcal{C}_{j,j'}$ at decision node $\{j, j'\}$. $D_{j,j'} \in \mathcal{D}$, the set of possible educational transition decisions taken by an individual over the life cycle. We assume that the environment is time-stationary and that educational decisions are irreversible. Each choice set contains two options: (a) remain at $j \in \mathcal{C}_{j,j'}$ or (b) continue on to $j' \in \mathcal{C}_{j,j'}$, where $j' \neq j$.

$Q_{j,j'} = 1$ denotes that a person gets to decision node $\{j, j'\}$. $Q_{j,j'} = 0$ if the person never visits decision node $\{j, j'\}$. The history of nodes visited by an agent can be described by the collection of $Q_{j,j'}$ such that $Q_{j,j'} = 1$. Clearly, $D_{j,j'}$ is not defined if $Q_{j,j'} = 0$. Formally we assign the value "0" to such undefined states, but we could assign any finite value.

We adopt the convention that $j = 0$ is the state of being without a high school credential; $j = 1$ denotes being a high school graduate; $j = 2$ denotes getting a GED (an option for dropouts); $j = 3$ denotes attending college; $j = 4$ denotes graduating college. The "j" denotes possible states a person can visit. We let s denote the realized final schooling level, and S denote the discrete random variable. A person who drops out of high school ($D_{0,1} = 1$) and does not earn a GED ($D_{0,2} = 1$) is a permanent dropout with $s = 0$. We observe post-schooling outcomes associated with each level of final educational attainment. In our sample, we have so few GEDs who attempt college that we ignore this possibility in our empirical analysis.[15]

Figure 2: Sequential Schooling Decisions



	Decision Taken:	
Decision Node $Q_{j,j'}$	$D_{j,j'} = 1$	$D_{j,j'} = 0$
$\{0, 1\}$	Graduate High School ($j = 1$)	Drop out of High School ($j = 0$)
$\{0, 2\}$	Get GED ($j = 2$)	High School Dropout ($j = 0$)
$\{1, 3\}$	Attend College ($j = 3$)	High School Graduate ($j = 1$)
$\{3, 4\}$	Graduate 4-yr college ($j = 4$)	Some College ($j = 3$)

[15] See Heckman, Humphries, and Kautz (2014).

2.1 A Sequential Model of Educational Attainment

Under general conditions, the optimal decision at each node is characterized by an index threshold-crossing model

$$D_{j,j'} = \begin{cases} 1 & \text{if } I_{j,j'} \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $I_{j,j'}$ is the perceived value (by the agent) of attaining schooling level j' for a person currently in educational state j. We do not take a position on the precise information set available to agents or the exact decision rule used. In principle, agents can make irrational choices or their educational choices could be governed by behavioral anomalies.

Associated with each final schooling state are a set of k potential outcomes for health, healthy behavior, and labor market outcomes. Define $Y^*_{k,s}$ as latent variables that map into potential outcomes $Y_{k,s}$:

$$Y_{k,s} = \begin{cases} Y^*_{k,s} & \text{if } Y_{k,s} \text{ is continuous,} \\ \mathbf{1}(Y^*_{k,s} \geq 0) & \text{if } Y_{k,s} \text{ is a binary outcome,} \end{cases} \tag{2}$$
$$k \in \{1, \ldots, K\}.$$

Let $H_s = 1$ if s is the highest level of attained schooling. $H_s = 0$ otherwise. Using the familiar switching regression framework of Quandt (1958), the observed outcome Y_k is

$$Y_k = \sum_{s \in S} H_s Y_{k,s}. \tag{3}$$

2.2 Parameterizations of the Decision Roles and Potential Outcomes

Following a well-established tradition in the literature,[16] we approximate $I_{j,j'}$ using a linear-in-the-parameters model:

$$I_{j,j'} = \boldsymbol{X}_{j,j'}\boldsymbol{\beta}_{j,j'} + \boldsymbol{\theta}\boldsymbol{\alpha}_{j,j'} - \nu_{j,j'}, \tag{4}$$

where $\boldsymbol{X}_{j,j'}$ is a vector of variables (and functions of these variables) observed by the economist that determine the schooling transition decision of the agent with schooling level j, $\boldsymbol{\theta}$ is a vector of unobserved (by the economist) endowments. This approximation is a starting point for a more general analysis of dynamic discrete choice models. Endowments $\boldsymbol{\theta}$ are not directly observed by the econometrician but are proxied by measures. $\boldsymbol{\theta}$ plays an important role in our model. Along with the observed variables, it generates dependence among schooling choices and outcomes. $\nu_{j,j'}$ represents an idiosyncratic error term assumed to be independent across agents and states. It plays the role of a random shock: $\nu_{j,j'} \perp\!\!\!\perp (\boldsymbol{X}_{j,j'}, \boldsymbol{\theta})$, where "$\perp\!\!\!\perp$" denotes statistical independence.

Latent variables generating outcomes are also approximated by a linear-in-the-parameters model.

$$Y^*_{k,s} = \boldsymbol{X}_{k,s}\boldsymbol{\beta}_{k,s} + \boldsymbol{\theta}\boldsymbol{\alpha}_{k,s} + \nu_{k,s}, \tag{5}$$

where $\boldsymbol{X}_{k,s}$ is a vector of observed controls relevant for outcome k and $\boldsymbol{\theta}$ is the vector of unobserved endowments. $\nu_{k,s}$ represents an idiosyncratic error term that satisfies $\nu_{k,s} \perp\!\!\!\perp (\boldsymbol{X}_{k,s}, \boldsymbol{\theta})$.

2.3 Measurement System for Unobserved Endowments θ

Most of the literature estimating the causal effect of schooling develops strategies for eliminating the effect of $\boldsymbol{\theta}$ in producing spurious relationships between schooling and

[16] See Heckman (1981), Cameron and Heckman (1987, 2001), Eckstein and Wolpin (1989), Geweke and Keane (2001), and Arcidiacono and Ellickson (2011).

outcomes.[17] Our approach is different. We proxy $\boldsymbol{\theta}$ to identify the interpretable sources of omitted variable bias and to determine how the unobservables mediate the causal effects of education. We follow a recent literature documenting the importance of both cognitive and noncognitive skills in shaping schooling choices and mediating the effects of schooling on outcomes.

Given $\boldsymbol{\theta}$, and conditional on $\boldsymbol{X}$, all educational choices and outcomes are assumed to be statistically independent. If $\boldsymbol{\theta}$ were observed, we could condition on $(\boldsymbol{\theta}, \boldsymbol{X})$ and achieve selection-bias-free estimates of causal effects and model parameters. This would be equivalent to a parametric version of matching on $\boldsymbol{\theta}$ and $\boldsymbol{X}$.[18] Both matching and the procedure in the paper assume that conditional on $\boldsymbol{\theta}$ and $\boldsymbol{X}$, outcomes and choices are statistically independent. In this paper, we do not directly measure $\boldsymbol{\theta}$. Instead, we proxy it and correct for the effects of measurement error on the proxy. Our analysis can be thought of as a parametric version of matching on mismeasured variables where we estimate and correct for the measurement error in the matching variables.[19] We test the robustness of our approach by allowing for an additional unproxied unobservable that accounts for dependence between schooling and economic outcomes not captured by our proxies. These additional sources of dependence can be identified without proxy measurements for them under the conditions stated in Heckman and Navarro (2007).

Following Carneiro, Hansen, and Heckman (2003) and Heckman, Stixrud, and Urzua (2006), we adjoin a system of measurement equations to proxy $\boldsymbol{\theta}$. We have access to information on cognitive and socioemotional measures. We thus link our paper to an emerging literature on the importance of cognitive and noncognitive skills in shaping schooling choices and outcomes.[20] The recent literature establishes that both cognitive and noncognitive skills can be shaped by interventions and that they are effective margins for social policy (see Heckman and Mosso, 2014, Heckman, Pinto, and Savelyev,

[17]See Heckman (2008).

[18]Matching is a version of selection on observables. See Carneiro, Hansen, and Heckman, 2003, and Abbring and Heckman, 2007. See also Heckman and Vytlacil (2007b).

[19]See Heckman, Pinto, and Savelyev (2013) and Conti, Heckman, Pinger, and Zanolini (2009) for applications of this approach.

[20]See, e.g., Borghans, Duckworth, Heckman, and ter Weel (2008); Almlund, Duckworth, Heckman, and Kautz (2011); Heckman, Humphries, and Kautz (2014).

2013).

Let θ^C and θ^{SE} denote the levels of cognitive and socioemotional endowments and suppose $\boldsymbol{\theta} = (\theta^C, \theta^{SE})$. We allow θ^C and θ^{SE} to be correlated. Let $T^C_{s,l}$ be the l^{th} cognitive test score, $T^{SE}_{s,l}$ the l^{th} socioemotional measure, and $T^{C,SE}_{s,l}$ the l^{th} measure influenced by both cognitive and socioemotional endowments, all measured at schooling level s. Parallel to the treatment of the index and outcome equations, we assume linear measurement systems for these variables:

$$T^C_{s,l} = \boldsymbol{X}^C_{s,l}\boldsymbol{\beta}^C_{s,l} + \theta^C \alpha^C_{s,l} + e^C_{s,l}, \tag{6}$$

$$T^{SE}_{s,l} = \boldsymbol{X}^{SE}_{s,l}\boldsymbol{\beta}^{SE}_{s,l} + \theta^{SE} \alpha^{SE}_{s,l} + e^{SE}_{s,l}, \tag{7}$$

$$T^{C,SE}_{s,l} = \boldsymbol{X}^{C,SE}_{s,l}\boldsymbol{\beta}^{C,SE}_{s,l} + \theta^C \tilde{\alpha}^C_{s,l} + \theta^{SE} \tilde{\alpha}^{SE}_{s,l} + e^{C,SE}_{s,l}. \tag{8}$$

The structure assumed in Equations (6), (7), and (8) is identified even when allowing for correlated factors, if we have one measure that is a determinant of cognitive endowments ($T^C_{s,l}$), one measure that is a determinant of socioemotional endowments ($T^{SE}_{s,l}$), at least three measures that load on both cognitive ability and socioemotional ability, and conventional normalizations are assumed.[21] We collect our assumptions about the dependence structure among the model unobservables in Table 1. In Section I of the Web Appendix, we test if additional unobservables beyond θ^C and θ^{SE} are required to capture the dependence between schooling and outcomes beyond that arising from observables. Our empirical estimates are essentially unchanged when we introduce a third factor to capture dependencies between schooling and outcomes not captured by the proxy factors. To simplify the exposition, in the main text we report results from models that use measurements to proxy $\boldsymbol{\theta}$.

[21] See, e.g., the discussion in Anderson and Rubin (1956) and Williams (2011). One of the factor loadings for θ^C and θ^{SE} has to be normalized to set the scale of the factors. Nonparametric identification of the distribution of $\boldsymbol{\theta}$ is justified by an appeal to the results in Cunha, Heckman, and Schennach (2010).

Table 1: Assumptions About Unobservables

Choice equation (2):	$\nu_{j,j'} \perp\!\!\!\perp \boldsymbol{X}_{k,l}\ \forall\, j, j', k, l$
	$\nu_{j,j'} \perp\!\!\!\perp \nu_{k,l} \quad \forall\,(k,l) \neq (j,j')$
Labor market and health outcomes (4) and (6):	
	$\nu_{k,s} \perp\!\!\!\perp \boldsymbol{X}_{k,s'} \quad \forall\, k, s, s'$
	$\nu_{k,s} \perp\!\!\!\perp \nu_{k,s'} \quad \forall\, s' \neq s, k$
Measurement system (9), (10), and (11):	
	$e^{q}_{s,l} \perp\!\!\!\perp \boldsymbol{X}^{q}_{s',l'} \quad \forall\, s, l, s', l',\ q \in \{C, SE, (C, SE)\}$
	$e^{q}_{s,l} \perp\!\!\!\perp e^{q'}_{s',l'}, \quad \forall\,(s',l',q') \neq (s,l,q),\ (q,q') \in \{C, SE, (C, SE)\}$
Cross-systems dependence:	
	$\boldsymbol{\theta} \perp\!\!\!\perp \left(\nu_{j,j'}, \boldsymbol{X}_{j,j'}, \nu_{k,s}, \boldsymbol{X}_{k,s}, e^{q}_{s,l}, \boldsymbol{X}^{q}_{s,l}\right) \quad \forall\, j, j', k, s, l,\ q \in \{C, SE, (C, SE)\}$
Mutual independence of errors across systems:	
	$\nu_{j,j'} \perp\!\!\!\perp \left(\boldsymbol{X}_{k,s}, \boldsymbol{X}^{q}_{s,l}\right) \quad \forall\, j, j', k, s, l,\ q \in \{C, SE, (C, SE)\}$
	$\nu_{k,s} \perp\!\!\!\perp \left(\boldsymbol{X}_{j,j'}, \boldsymbol{X}^{q}_{s',l}\right) \quad \forall\, j, j', k, s, l, s',\ q \in \{C, SE, (C, SE)\}$
	$e^{q}_{s,l} \perp\!\!\!\perp \left(\boldsymbol{X}_{j,j'}, \boldsymbol{X}_{k,s'}\right) \quad \forall\, j, j', k, s, l, s',\ q \in \{C, SE, (C, SE)\}$
	$\nu_{j,j'} \perp\!\!\!\perp e^{q}_{s,l} \quad \forall\, j, j', s, l,\ q \in \{C, SE, (C, SE)\}$
	$\nu_{k,s} \perp\!\!\!\perp \left(\nu_{j,j'}, e^{q}_{s',l}\right) \quad \forall\, j, j', s, k, s', l,\ q \in \{C, SE, (C, SE)\}$

Note: For linear models the independence assumption can be relaxed to allow the error terms to share a common component (e.g. ε_k) across schooling levels (i.e. instead assume $\hat{\nu}_{k,s} \perp\!\!\!\perp \hat{\nu}_{k,s'}$, where $\nu_{k,s} = \varepsilon_k + \hat{\nu}_{k,s}$).

2.4 Sources of Identification

Our model has multiple sources of identification. First, if $\boldsymbol{\theta}$ were measured without error, the model would be identified by conditioning of $\boldsymbol{\theta}$ (a version of matching on $\boldsymbol{X}$ and $\boldsymbol{\theta}$ in a parametric model). If it is measured with error but all components are proxied, and the identifying restrictions for the factor models given in the previous section are satisfied, we can use the extension for matching on mismeasured variables developed in Carneiro, Hansen, and Heckman (2003) and Heckman, Pinto, and Savelyev (2013). Under either set of conditions, the model is identified without making any distributional assumptions on the unobservables (see, e.g., Cunha, Heckman, and Schennach, 2010). We also have access to transition-specific instruments, variables in $\boldsymbol{X}_{j,j'}$, not in $\boldsymbol{X}_{k,s}$, assumed to be independent of the model unobservables. The benefit of access to instrumental variables is that they allow us to test the validity of either version of the matching assumption. Under support conditions on the instruments specified in Heckman and Navarro (2007), the model is identified without invoking any distributional assumptions on the unobservables. We approximate the distribution of unobservables using mixtures of normal sieve estimators (see Chen, 2007).

3 Defining Treatment Effects

Under our assumptions, the model estimates distributions of counterfactual outcomes. Hence a variety of treatment effects for the effect of education on labor market and health outcomes can be generated from it. They can be used to predict the effects of manipulating education levels through different channels for people of different ability levels. They allow us to understand the effectiveness of policy for different segments of the population.

We consider two different formulations of treatment effects. The first compares returns between two terminal schooling levels. The second estimates the treatment effect of specific educational decisions, inclusive of the continuation values associated with future decisions.

The traditional literature on estimating the returns to schooling defines its parameters in terms of the returns generated from going from one final schooling level to another (Becker, 1964). It ignores the sequential nature of schooling and the options created by going to an additional level of school. For example, after graduating from high school, an agent may enroll in college. After enrolling, the agent may choose to earn a four-year degree. The benefits of graduating from high school include the options which subsequent education makes possible. (See Weisbrod, 1962; Comay, Melnik, and Pollatschek, 1973; Altonji, 1993 and Cameron and Heckman, 1993.)

Treatment effects can be identified at each node in the educational choice tree of Figure 2. For example, we estimate the treatment effect for deciding to graduate from high school or drop out (node $\{0, 1\}$). Once agents graduate from high school, they have the option of going to college and even graduating from college. Similarly, once agents drop out, they have the option of getting a GED. The full returns to early choices include the benefits from access to additional educational options.

3.1 Traditional Treatment Effects: Differences Across Final Schooling Levels

We estimate the traditional returns to education defined as the gains from choosing between terminal schooling levels. Let $Y_{s'}$ be an outcome at schooling level s' and Y_s be an outcome at schooling level s. Conditioning on $\boldsymbol{X} = x$ and $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$, the average treatment effect of s compared to s' is $E(Y_s - Y_{s'}|X = x, \theta = \overline{\theta})$. Measured over the entire population it is

$$ATE^{*}_{s,s'} \equiv \iint E(Y_s - Y_{s'}|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{x}, \overline{\boldsymbol{\theta}}). \tag{9}$$

The average treatment effect calculated by averaging over the subset of the population that completes one of the two final schooling levels is

$$ATE_{s,s'} \equiv \iint E(Y_s - Y_{s'}|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{x}, \overline{\boldsymbol{\theta}} \,|\, H_s + H_{s'} = 1). \tag{10}$$

3.2 Dynamic Treatment Effects

We also estimate treatment effects associated with each decision node. These take into account the benefits associated with the options opened up by educational choices. This treatment effect is the difference in expected outcomes arising from changing a single educational decision in a sequential schooling model and tracing through its consequences. We estimate the continuation value as the probability-weighted benefit of further educational choices using probabilities perceived by the agent. In computing a version of these probabilities to identify treatment effects (but only for this purpose), we assume rational expectations: the empirical probabilities are assumed to be what the agent acts on.[22] The expected value associated with *fixing* a particular education

[22]We do *not* impose rational expectations in estimating the choice model, just in interpreting it.

transition ($D_{j,j'} = 1$) for an individual with $\boldsymbol{X} = \boldsymbol{X}$ and $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ is

$$\begin{aligned} & E\left(Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{j,j'} = 1\right) \\ & \equiv \sum_{s \in \mathcal{S}} Pr\left(s|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{j,j'} = 1\right) \times E\left(Y_s|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}\right). \end{aligned}$$ [23]

The expectation (E) on the left-hand side is over future educational choices and idiosyncratic shocks. $Pr\left(s|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{j,j'} = 1\right)$ is the probability that the individual stops at education level s when fixing $D_{j,j'} = 1$. Y_s is the value of the outcome if the individual stops at education level s. For example, the choice of graduating from high school opens up the possibility of enrolling in college and possibly graduating from college.[24]

The person-specific treatment effect for an individual making a decision at node (j, j') deciding between going on to j' or stopping at j is the difference between the expected value of the two decisions:

$$\begin{aligned} T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}] \equiv & E(Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ Q_{j,j'} = 1, Fix\ D_{j,j'} = 1) \\ & - E(Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ Q_{j,j'} = 1, Fix\ D_{j,j'} = 0). \end{aligned}$$

The person-specific treatment effect not only takes into account the direct effect of the decision, but also includes the value of any additional schooling. Averaged over the full population

$$ATE^{*D}_{j,j'} \equiv \iint T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}]\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{x}, \overline{\boldsymbol{\theta}}). \quad (11)$$

The corresponding parameter for those who ever visit the decision node $\{j, j'\}$ is

$$ATE^{D}_{j,j'} \equiv \iint T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}]\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{x}, \overline{\boldsymbol{\theta}} \,|\, Q_{j,j'} = 1). \quad (12)$$

[23]The distinction between *fixing* and *conditioning* traces back to Haavelmo (1943). For a recent analysis see Heckman and Pinto (2013). Under the assumptions in Table 1, fixing and conditioning produce the same causal parameter conditioning on $\boldsymbol{X}$ and $\boldsymbol{\theta}$.

[24]The expected outcome for an individual with characteristics $\boldsymbol{X}$ and $\boldsymbol{\theta}$ who chooses to graduate from high school ($D_{0,1} = 1$) is thus $E(Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{0,1} = 1) = Pr(s = 1|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{0,1} = 1) \times E(Y_1, |\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) + Pr(s = 3|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{0,1} = 1) \times E(Y_3|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) + Pr(s = 4|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{0,1} = 1) \times E(Y_4|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})$.

The average marginal treatment effect is the average effect of choosing an additional level of schooling for individuals who are at the margin of indifference between the two choices:

$$AMTE_{j,j'} \equiv \iint T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}]\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(x, \bar{\theta} \mid |I_{j,j'}| \leq \varepsilon), \tag{13}$$

where ε is an arbitrarily small neighborhood around the margin of indifference. AMTE defines causal effects at well-defined and empirically identified margins of choice. This is the proper measure of marginal gross benefits for evaluating social policies.[25]

Each treatment effect can be decomposed into direct effects and a continuation value. For example, the continuation value of graduating from high school is the probability that the individual enrolls in college times the expected wage benefit of having some college plus the probability of completing college times the wage benefit of completing college and stops there:[26]

$$\begin{aligned} & CV_{0,1}(Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \\ & \qquad = E(Y_4 - Y_1|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \times \Pr(s = 4|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{1,0} = 1) \\ & \qquad + E(Y_3 - Y_1|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \times \Pr(s = 3|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{1,0} = 1). \end{aligned}$$

We report estimates of the continuation value ratio (CVR) to summarize the relative importance of the CV. It is the average continuation value (ACV) divided by the average treatment effect for the population considered ($CVR = \frac{ACV}{ATE}$).

3.3 Policy Relevant Treatment Effects

The policy relevant treatment effect (PRTE) is the average treatment effect for those induced to change their educational choices in response to a particular policy intervention. Let Y^p be the aggregate outcome under policy p. Let $S(p)$ be the schooling selected

[25]See, e.g., Heckman and Vytlacil (2007a) and Carneiro, Heckman, and Vytlacil (2010).

[26]The direct effect of graduating from high school is $DTE_{0,1}(Y|\boldsymbol{X} = \boldsymbol{x},\, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) = E(Y_1 - Y_0|\boldsymbol{X} = \boldsymbol{x_1}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) - E(Y_2 - Y_0|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \times \Pr(s = 2|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{0,1} = 0)$. The second term on the right-hand side arises from the forgone option of taking the GED.

under policy p. The policy relevant treatment effect from implementing policy p compared to policy p' is:

$$PRTE_{p,p'} \equiv \iint E(Y^p - Y^{p'}|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}), dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{X}, \overline{\boldsymbol{\theta}}|S(p) \neq S(p')), \qquad (14)$$

where $S(p) \neq S(p')$ denotes the set of people and their associated $\boldsymbol{\theta}$, $\boldsymbol{X}$ values for whom attained schooling levels differ under the two policies.

4 Data

We use the 1979 National Longitudinal Survey of Youth (NLSY79) to estimate our model. It is a nationally representative sample of men and women born in the years 1957–1964. Respondents were first interviewed in 1979 when they were 14–22 years of age. The NLSY surveyed its participants annually from 1979 to 1992 and biennially since 1992. The NLSY measures a variety of adult outcomes including income and health. The survey also measures many other aspects of the respondents' lives, such as educational attainment, fertility, scores on achievement tests, high school grades, and family background variables. This paper uses the core sample of males, which, after removing observations with missing covariates, contains 2242 individuals.[27] We report results for samples that pool race groups.

4.1 Outcomes, Transitions, and Final Attainment Levels

This paper considers the effect of education on three different health and health-related outcomes: overall physical health, smoking, and self-esteem.[28] As a measure of physical

[27]Respondents were dropped from the analysis if they did not have valid ASVAB scores, missed multiple rounds of interview, had implausible educational histories, were missing control variables which could not be imputed, or had implausible labor market histories. A number of imputations where made as necessary. Previous years' covariates were used when covariates where not available for a needed year (such as region of residence). Responses from adjacent years were used for some outcomes when outcome variables were missing at the age of interest. Mother's education and father's education were imputed when missing. See Web Appendix Section A for the analysis of the deleted observations.

[28]The literature focuses primary attention on the effect of mortality and on smoking. See Cutler and Lleras-Muney (2010).

health, we use the PCS-12 scale, the Physical Component Summary obtained from the SF-12.[29] The SF-12 in turn is designed to provide a measure of the respondent's mental and physical health irrespective of their proclivity to use formal health services. We also study smoking at age 30 as an additional measure of healthy behaviors. It is a self-reported, binary variable recording whether the individual smoked daily at age 30. As a measure of mental health, the effect of education on self-esteem is considered. Self esteem is measured using Rosenberg's self-esteem scale collected in 2006, when individuals were in their 40s.

We also analyze the effect of education on log wages at age 30 as a traditional benchmark. Details on the construction of these outcome variables are presented in the Web Appendix Section A.

We estimate models with the four different transitions and five final schooling levels depicted in Figure 2. Education at age 30 is treated as the respondent's final schooling level.[30]

4.2 Measurement System (T)

Our approach to estimating the impact of cognitive and socioemotional measures on schooling choices and outcomes improves on the traditional approach, which uses indices of direct measures of behavior, designated as cognitive and socioemotional indices, to proxy latent traits. We allow for measurement error and use factor analysis to let the data determine the weights on specific measures used in forming indices.[31]

As noted by Almlund, Duckworth, Heckman, and Kautz (2011) and Heckman and Kautz (2012, 2014), a fundamental identification problem plagues the extraction of psychological characteristics. Traits are measured from behaviors that can also be affected by incentives and other traits. Even after controlling for these incentives and other traits, some normalizations are necessary to operationalize the measures of traits,

[29] SF-12 is a 12-question health survey designed by John Ware of the New England Medical Center Hospital (see Ware, Kosinski, and Keller (1996) and Gandek, Ware, Aaronson, Apolone, Bjorner, Brazier, Bullinger, Kaasa, Leplege, Prieto, and Sullivan (1998).

[30] A negligible fraction of individuals change schooling levels after age 30.

[31] See Cunha and Heckman (2008).

and distinguish one trait from another. Even if this distinction cannot be made, we can condition on the (entire) set of traits without distinguishing which particular traits produce outcomes. Hence, our estimates of the causal effects of schooling on outcomes do *not* require that we solve these identification problems. Our estimates of effects of specific factors do.

We identify the cognitive and socioemotional factors used in this paper from an auxiliary measurement system fit jointly with a model of educational choice. Following Hansen, Heckman, and Mullen (2004), we control for the effect of schooling at the time the measurements are taken on the measurements to control for feedback from schooling to measured traits. We do not use the outcome measures $(\boldsymbol{Y})$ in extracting the distributions of latent factors to avoid getting tautologically good fits between outcomes and factors and to render the factors interpretable.

Sub-tests from the Armed Services Vocational Aptitude Battery (ASVAB) are used as dedicated measures of cognitive ability and are assumed not to be determined by the socioemotional factor. Specifically, we consider scores from Arithmetic Reasoning, Coding Speed, Paragraph Comprehension, Word Knowledge, Math Knowledge, and Numerical Operations.[32]

Academic success (measured by GPA) depends on cognitive ability, but also depends strongly on socioemotional traits such as conscientiousness, self-control, and self-discipline. This motivates our identification strategy of including both a cognitive and socioemotional factor as determinants of 9th grade GPA, as much of the variance in this measure not explained through cognitive test scores has been shown to be related to socioemotional traits.[33]

To identify the socioemotional factor, we use measures of participation in minor

[32] A subset of these tests is used to construct the Armed Forces Qualification Test (AFQT) score, which is commonly used as a measure of cognitive ability. AFQT scores are often interpreted as proxies for cognitive ability (Herrnstein and Murray, 1994). See the discussion in Almlund, Duckworth, Heckman, and Kautz (2011).

[33] As noted by Borghans, Golsteyn, Heckman, and Humphries (2011) and Almlund, Duckworth, Heckman, and Kautz (2011), the principal determinants of the grade point average are personality traits and not cognition. Similarly, Duckworth and Seligman (2005) find that self-discipline predicts GPA in 8th graders better than IQ. See also Duckworth, Quinn, and Tsukayama (2010).

risky or reckless activity in 1979 in the measurement system for the socioemotional endowment.[34] In order to identify the distribution of correlated factors, risky behavior is restricted to not load on the cognitive factor.[35]

As a robustness check, we include five additional measures of risky adolescent behavior to check our estimates based on the non-cognitive factor.[36] We consider violent behavior in 1979 (fighting at school or work and hitting or threatening to hit someone), tried marijuana before age 15, daily smoking before age 15, regular drinking before age 15, and any intercourse before age 15. For violent behavior, we control for the potential effect of schooling on the outcome. The estimates based on including these measures are essentially the same as the ones reported in the text.

4.3 Control Variables (X)

For every outcome, measure, and educational choice, we control for race, broken home status, number of siblings, mother's education, father's education, and family income in 1979. We additionally control for region of residence and urban status at the time the relevant measure, decision, or outcome was assessed.[37] For log wages at age 30, we additionally control for local economic conditions at age 30.

The models for educational choice include additional choice-specific covariates. Following Carneiro, Heckman, and Vytlacil (2011), we control for both long-run economic conditions measured by unemployment and current deviations from those conditions. By controlling for the long-run local economic environment, local unemployment variations capture current economic shocks that might affect schooling decisions.

[34]Preliminary data analysis suggested that one measure of risky behavior is the least correlated cognitive endowments among our measures of socioemotional traits. This variable is a binary variable which is unity if an agent answers yes to any of the following questions in 1980: "Taken something from the store without paying for it," "Purposely destroyed or damaged property that did not belong to you?," "Other than from a store, taken something that did not belong to you worth under \$50?," and "Tried to get something by lying to a person about what you would do for him, that is, tried to con someone?"

[35]These measures are used for estimating one of the three-factor models discussed in Web Appendix Section I.

[36]Gullone and Moore (2000) present a line of research which studies the relationship between personality traits and adolescent risk-behavior.

[37]Based on the data, we assume that high school, GED certification, and college enrollment decisions occur at age 17 while the choice to graduate from college is made at age 22.

4.4 Exclusion Restrictions

Identification of our model does not rest solely on the conditional independence assumptions listed in Table 1 which, by themselves, could justify identification of our model as a version of matching on mismeasured variables. Under sufficient support conditions on the regressors, the model is nonparametrically identified and does not require either conditional independence assumptions or normality assumptions.[38] To nonparametrically identify treatment effects without invoking the full set of conditional independence assumptions, node-specific instruments are critical. We have a variety of exclusion restrictions that affect choices but not outcomes. These are listed in the bottom five rows of Table 2. The node-specific exclusion restrictions are noted at the base of the table.

In a dynamic forward-looking model, instruments at later stages that are known at earlier stages cannot be excluded from choices at earlier stages. Following the literature on rational expectations econometrics (see Hansen and Sargent, 1980), stage-specific expectations of the future-stage realized value of instruments qualify as valid instruments. We find that candidate instruments based on future variables (e.g., college tuition for high school graduation) were not statistically significant predictors of early-stage (high school graduation) decisions. In our samples, future realizations of the potential instruments do not predict previous stage-specific schooling choices. Invoking linearity of the effect of schooling on outcomes widely used in the literature (Card, 1999, 2001) avoids the need for stage-specific instruments. However, linearity in years of schooling for wage equations is decisively rejected in many data sets.[39] In addition, different schooling levels are qualitatively different.[40]

[38]Heckman and Navarro (2007) and Abbring and Heckman (2007).

[39]See the evidence discussed in Heckman, Lochner, and Todd (2006). See Web Appendix Section O for our evidence on nonlinearity.

[40]For example, there is no specific number of years of schooling to assign to GEDs.

Table 2: Control Variables and Instruments Used in the Analysis

Variables	Measurement Equations	Choice	Outcomes
Race	x	x	x
Broken Home	x	x	x
Number of Siblings	x	x	x
Parents' Education	x	x	x
Family Income (1979)	x	x	x
Region of Residence	x	x	x
Urban Status	x	x	x
Age[a]	x	x	
Local Unemployment[b]			x
Local Long-Run Unemployment		x	
Instruments			
Local Unemployment at Age 17[c]		x	
Local Unemployment at Age 22[d]		x	
GED Test Difficulty[e]		x	
Local College Tuition at Age 17[f]		x	
Local College Tuition at Age 22[g]		x	

Notes: [a] Age in 1979 is included as a cohort control. We also included individual cohort dummies which did not change the results. [b] For economic outcomes, local unemployment at the time of the outcome. [c] This is an instrument for choices at nodes $\{0,1\}$, $\{0,2\}$, and $\{1,3\}$. [d]This is an instrument for the choice at $\{3,4\}$. Region and urban dummies are specific to the age that the measurement, educational choice, or outcome occurred. [e]GED test difficulty only enters the decision to earn a GED. GED difficulty is proxied by the percent of high school graduates able to pass the test in one try given the state's chosen average and minimum score requirements. Control variables at age 17 are used for the high school graduation, GED certification, and college enrollment decisions. Control variables at age 22 are used for the college graduation decision. Control variables at age 14 are used for 9th grade GPA, and control variables from 1979 are used for the ASVAB tests. [f]Local college tuition at age 17 only enters the college enrollment graduation decisions. [g]Local college tuition at age 22 only enters the college completion equation.

5 Estimation Strategy

We estimate the model in two stages. The distribution of latent endowments and the model of schooling decisions are estimated in the first stage. The outcome equations are estimated in the second stage using estimates from the first stage.

We follow Hansen, Heckman, and Mullen (2004), and correct estimated factor distributions for the causal effect of schooling on the measurements by jointly estimating the schooling choice and measurement equations. The distribution of the latent factors is estimated using data on only educational choices and measurements. This allows us to interpret the factors as cognitive and socioemotional endowments. It links our estimates to an emerging literature on the economics of personality and psychological traits but is not strictly required if we only seek to control for selection in schooling choices. We do not use the final outcome system to estimate the factors, thus avoiding producing tautologically strong predictions from the estimated factors.

Assuming independence across individuals, the likelihood is:

$$\begin{aligned}\mathcal{L} &= \prod_i f(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{T}_i | \boldsymbol{X}_i) \\ &= \prod_i \int f(\boldsymbol{Y}_i | \boldsymbol{D}_i \boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{D}_i, \boldsymbol{T}_i | \boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta},\end{aligned}$$

where $f(\cdot)$ denotes a probability density function. The last step is justified from the assumptions listed in Table 1. For the first stage, the sample likelihood is

$$\mathcal{L}^1 = \prod_i \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} f(\boldsymbol{D}_i, \boldsymbol{T}_i | \boldsymbol{X}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \, dF_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}), \tag{15}$$

where we integrate over the distributions of the latent factors. We approximate the factor distribution using a mixture of normals.[41] We assume that the idiosyncratic shocks are mean zero normal variates. The goal of the first stage is to secure estimators of $f(\boldsymbol{D}_i, \boldsymbol{T}_i \mid \boldsymbol{X}_i, \boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$. In the second stage, we use first stage estimates (denoted

[41] Mixtures of normals can be used to identify the true density nonparametrically, where the number of mixtures can be increased based on the size of the sample. For a discussion of sieve estimators, see Chen (2007).

"ˆ") to form the likelihood

$$\mathcal{L}^2 = \prod_i \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} f(\boldsymbol{Y}_i | \boldsymbol{D}_i, \boldsymbol{X}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \hat{f}(\boldsymbol{D}_i, \boldsymbol{T}_i | \boldsymbol{X}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) d\hat{F}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}). \tag{16}$$

Since outcomes ($\boldsymbol{Y}_i$) are independent from the first stage outcomes conditional on $\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{D}_i$ and we impose no cross-equation restrictions, we obtain consistent estimates of the parameters for the adult outcomes. Each stage is estimated using maximum-likelihood. Standard errors and confidence intervals are calculated by estimating two hundred bootstrap samples for the combined stages.

6 Empirical Estimates

Since the model is nonlinear and multidimensional, in the main text we only report interpretable simulations of it.[42] We randomly draw regressors from the observations and the estimated factor distributions to simulate the various outcomes.

We present our empirical analysis in the following order. Section 6.1 discusses the estimates of the measurement system and the estimated effects of the endowments on schooling, labor market, and health outcomes. Section 6.2 presents estimates of the treatment effects of education. Section 6.3 gives estimates of the policy relevant treatment effect of a subsidy to college tuition. Section 6.4 discusses the importance of the latent cognitive and socioemotional endowments.

6.1 The measurement of endowments and their effects on outcomes

Figure 3 shows the variance decomposition of the measures. The latent factors explain between 20% and 70% of the variance of the ASVAB tests and school grades. While the socioemotional factor explains only 2%–3% of the variation in reckless behavior, it has a statistically significant effect on outcomes for those with at least a high school degree.[43]

[42]Parameter estimates for individual equations are reported in the Web Appendix.
[43]See Web Appendix Section C.

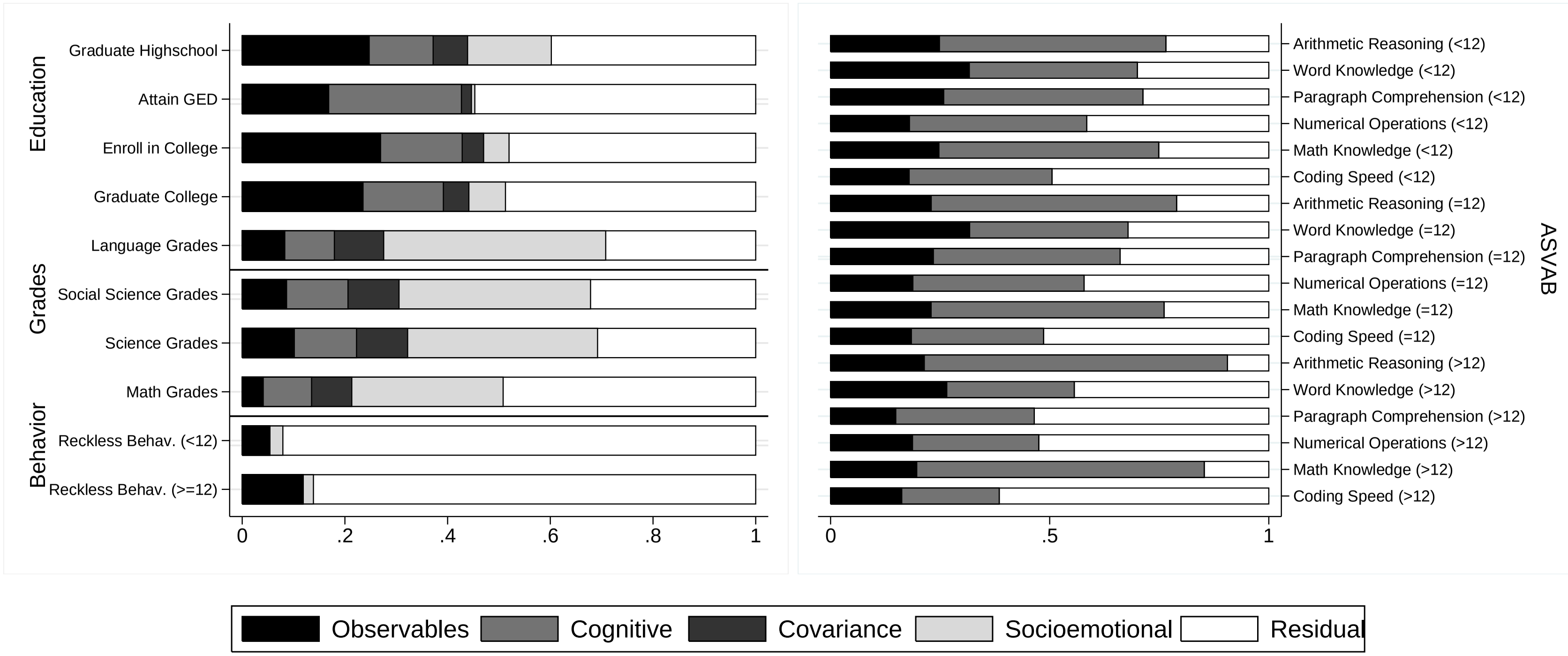
Although observed and latent variables combined explain a large part of the variance of the test scores and grades, there is a still significant amount of measurement error (labeled as "Residual" in the figure).[44] We test and reject the hypothesis that a bivariate normal model produces good model fit in favor of a model with a mixture of two normals for the factor distribution.[45] Simulations of the model show that the estimated model fits the data well.[46] We find a positive and statistically significant correlation between the cognitive and socioemotional/personality endowments ($\rho = 0.23$).

[44]This implies that any predictions of the factors would also have a significant amount of measurement error.

[45]The likelihood ratio test was used to test for the appropriate number of mixtures.

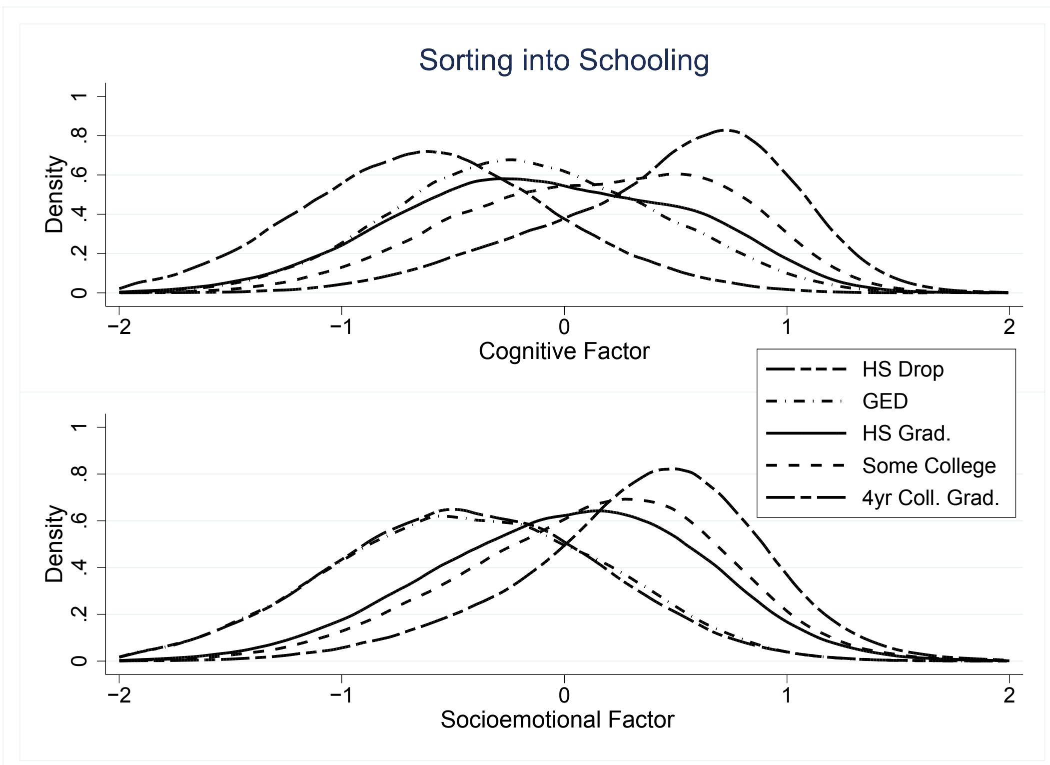
[46]The goodness of fit measurements are made for the various outcomes and measurement systems. Goodness of fit for discrete outcomes is tested using a χ^2 test of fit of the model to data. For continuous outcomes, the equality of the model and data are tested using t-tests. We test for mean equivalence of the means for many sub-populations and jointly test if the means are equivalent for all sub-populations using a χ^2 test. Tests of goodness of fit are found in Section E in the Web Appendix.

Figure 3: Decomposing Variances in the Measurement System



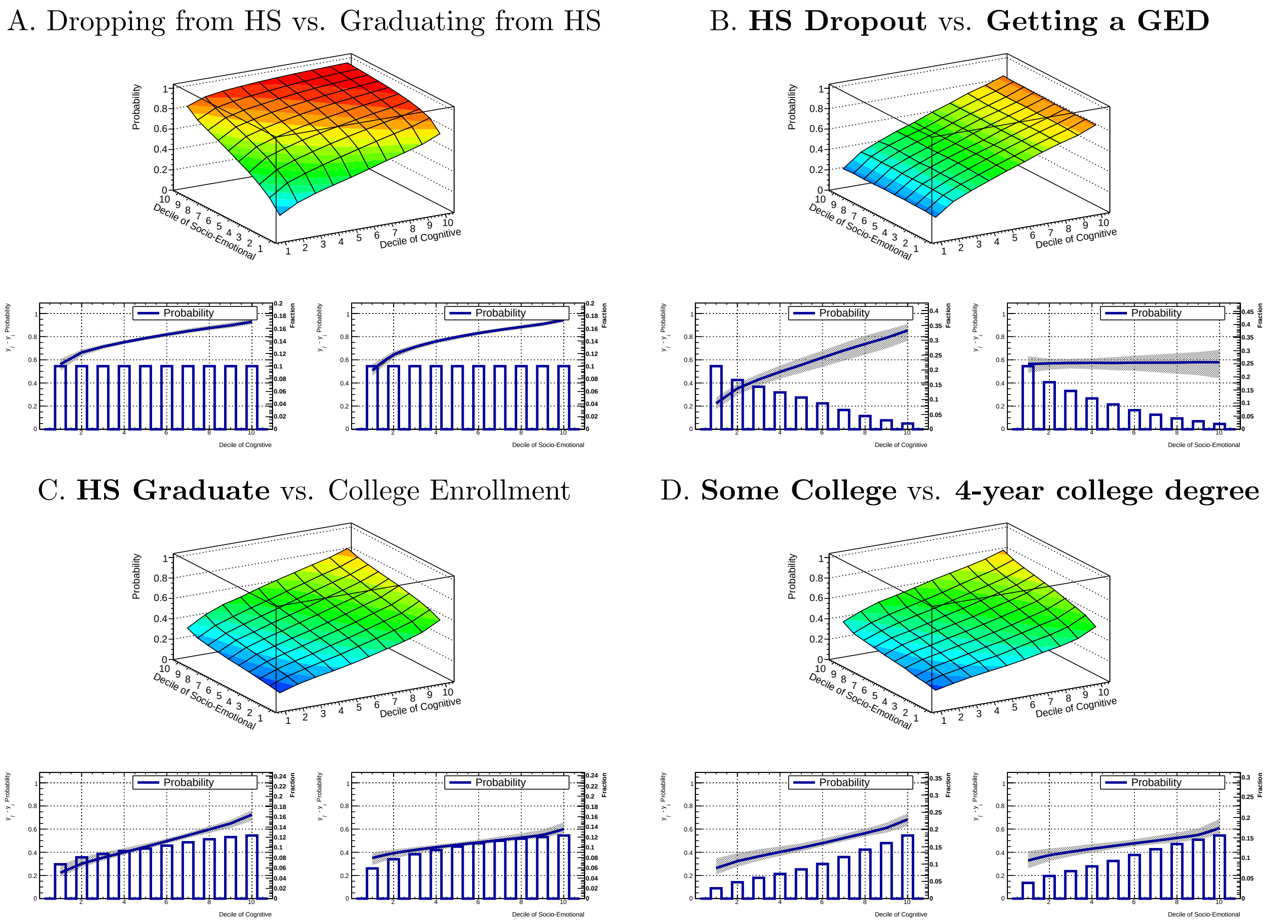
Notes: Bars indicate the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socioemotional factors (θ_C, θ_{SE}), and remaining unobservables (ϵ). For continuous outcomes we decompose the observed variance, while for discrete outcomes we decompose the variance of the latent index. Given the assumption that the factors, observable characteristic, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\boldsymbol{X'\beta}) + var(\boldsymbol{\theta'\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\boldsymbol{X'\beta}) + var(\boldsymbol{\theta'\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha'\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{NC}\alpha_{NC}) + var(\theta_{NC}\alpha_{NC})$. In the legend above, for continuous outcomes, "Observables" is $var(\boldsymbol{X'\beta})/var(Y)$, "Cognitive" is $var(\theta_C\alpha_C)/var(Y)$, "Covariance" is $2cov(\theta_C\alpha_C, \theta_{NC}\alpha_{NC})/var(Y)$, and "Socioemotional" is $var(\theta_{NC}\alpha_{NC})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalizes by $var(I)$ rather than $var(Y)$. The ASVAB tests are assumed to not depend on the socioemotional endowment, while reckless behavior is assumed to not depend on the cognitive endowment. The ASVAB and behavior models are estimated separately for those with less than twelve years (< 12), those who are high school graduates ($= 12$), and those who have attended college (> 12) at the time they took the ASVAB tests. Minor reckless behavior, which is also measured in 1979, also estimates models separately for those with less than 12 years and those with 12 or more years of schooling.

Figure 4: Distribution of factors by schooling level



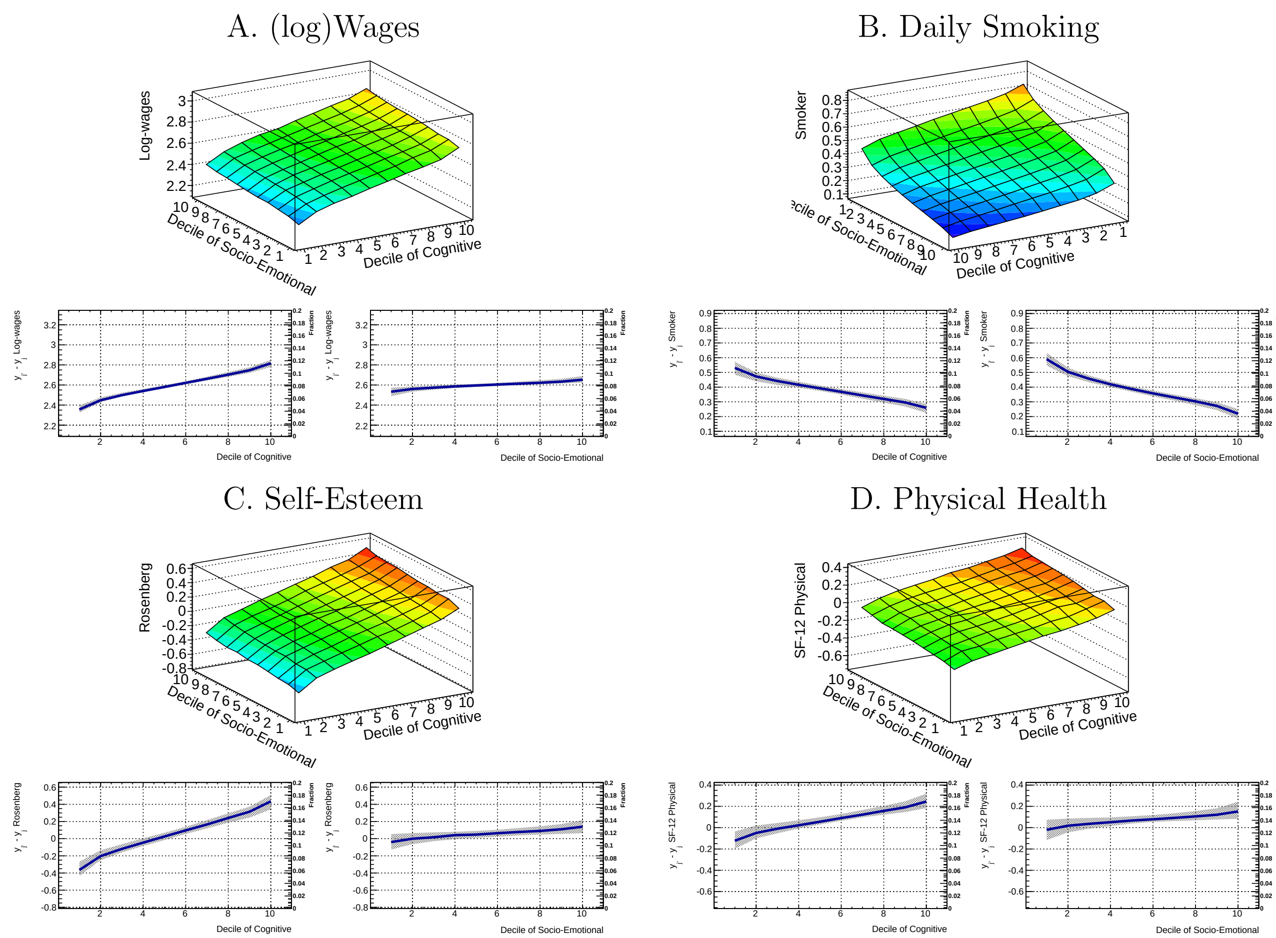
Note: The factors are simulated from the estimates of the model. The simulated data contain 1 million observations.

Figure 5: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters)



Notes: For each of the four educational choices, we present three figures that study the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels 0 and 1, the first subfigure (top) presents $Prob(D|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. $Prob(D|d^C, d^{SE})$ is computed for those who reach the decision node involving a decision between levels 0 and 1. The bottom left subfigures present $Prob(D|d^C)$ where the socioemotional factor is integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels 0 and 1. The bottom right subfigures present $Prob(D|d^{SE})$ for a given decile of socioemotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels 0 and 1.

Figure 6: The Effect of Cognitive and Socioemotional endowments



Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socioemotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socioemotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socioemotional endowments. Notice that we define as "decile 1" the decile with the lowest values of endowments and "decile 10" as the decile with the highest levels of endowments. The bottom left figure displays the average levels of endowment across deciles of cognitive endowments. The bottom right figure mimics the structure of the left-hand side figure but now for the socioemotional endowment.

There may be concern that the socioemotional factor is describing something like academic ability, since it is partly based on grades.[47] In order to address this issue, the role of the cognitive and socioemotional factors in adverse adolescent behaviors is investigated. Table 3 reports estimates for smoking, drinking, intercourse and violent behaviors. The factor loadings (the coefficients for "cognitive" and "socioemotional" factors at the base of each table) show that the socioemotional factor plays a significant role in these adverse behaviors, whereas the cognitive loadings are either statistically insignificant or much smaller than the socioemotional loadings.

[47]Our exclusion restriction for the socioemotional factor is "risky and reckless behavior in 1979", which is a binary measure of whether the person had ever done any of the following things: (1) purposefully damaged another person's property, (2) stolen an item of value from another person that was worth less than \$50, (3) stolen a small item from a store, or (4) tried to get something from someone by lying about what they would do in return. Not included as part of the measurement system, we have additional binary measures for violent behavior in 1979, daily smoking before age 15, tried marijuana before 15, regular drinking before 15, and sexual intercourse before age 15. These five measures were excluded from the measurement system as they are extreme enough that they may affect schooling decisions and later life health. For example, we did not want to predict later life health decisions with early life health decisions, or educational choice by actions that could lead to incarceration (such as violent behavior). We include these as outcomes that do not inform our measurement system to help us interpret our factor. We chose risky and reckless behavior in 1979 as our exclusion restriction because it was less likely to directly determine educational outcomes. All of the risky measures had low and statistically insignificant correlations with AFQT ($|\rho| < 0.05$). See section N.1.

Table 3: Early Outcomes: Estimates for "Early Risky Behaviors"

Variable	Tried Marijuana[a]		Daily Smoking[a]		Regular Drinking[a]		Intercourse[a]		Violent in 1979[b]			
									<12 yrs school		≥12 yrs school	
	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.
Black	-0.327	0.100	-0.331	0.112	-0.242	0.108	0.600	0.099	-0.253	0.125	0.154	0.157
Hispanic	-0.152	0.124	-0.504	0.150	-0.009	0.130	-0.033	0.140	-0.316	0.161	-0.054	0.204
Broken Home	0.444	0.072	0.411	0.081	0.243	0.077	0.366	0.080	0.220	0.094	0.130	0.116
Number of Siblings	0.024	0.014	0.035	0.015	0.027	0.015	0.011	0.016	0.009	0.018	0.013	0.021
Mother's Education	0.007	0.015	-0.020	0.017	-0.001	0.016	-0.022	0.017	0.030	0.019	-0.033	0.023
Father's Education	-0.007	0.011	-0.037	0.013	-0.003	0.012	-0.027	0.013	-0.034	0.015	0.009	0.016
Family Income	0.000	0.003	-0.002	0.003	-0.001	0.003	-0.003	0.003	-0.006	0.004	-0.006	0.003
Intercept	-0.770	0.185	-0.487	0.213	-1.059	0.198	-0.860	0.216	-3.223	3.597	15.948	10.354
Urban	0.258	0.072	0.116	0.081	0.095	0.077	0.212	0.087	0.131	0.093	0.012	0.104
South	-0.103	0.066	-0.027	0.075	0.066	0.071	0.104	0.076	-0.204	0.098	0.050	0.113
West									-0.125	0.117	0.149	0.132
Northeast									-0.184	0.120	0.053	0.122
Age									0.517	0.384	-1.434	0.990
Age^2									-0.017	0.010	0.033	0.024
College Attendance											-0.132	0.124
Cognitive	-0.113	0.048	-0.210	0.054	-0.142	0.051	-0.281	0.057	-0.160	0.062	-0.233	0.073
Socioemotional	-0.604	0.059	-0.526	0.064	-0.287	0.061	-0.401	0.066	-0.497	0.078	-0.262	0.090
N	2239		2176		2231		2218		1272		909	

The numbers in this table represent the estimated coefficients and standard errors associated with binary choice models of early risky behaviors on the set of controls presented in rows. Information about living in the West and Northeast is only available in 1979. [a] The dependent variable takes a value of one if the individual has reported the behavior before age 15, and zero otherwise. [b] The variable "Violent" takes a value of one if the individual participated in fighting or assault in 1979, and is estimated separately by education level to account for age differences in 1979. For violent behavior, which is measured in 1979 rather than at a specific age, we include age and age squared.

Table 4: Estimates for Schooling Choice Model

Variable	$D_{0,1}$: Graduate HS vs. Drop out of HS		$D_{0,2}$: **GED** vs. **HS Dropout**		$D_{1,3}$: Enroll College vs. **HS Graduate**		$D_{3,4}$: **4-year College Degree** vs. **Some College**	
	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.
Black	0.075	0.129	-0.119	0.178	0.174	0.140	0.010	0.196
Hispanic	0.649	0.179	-0.083	0.252	0.643	0.175	0.410	0.255
Broken Home	-0.484	0.101	-0.240	0.140	-0.047	0.103	-0.278	0.141
Number of Siblings	-0.048	0.019	0.003	0.027	-0.053	0.019	-0.028	0.027
Mother's Education	0.127	0.022	0.073	0.033	0.097	0.021	0.100	0.027
Father's Education	0.066	0.016	0.038	0.026	0.127	0.015	0.103	0.019
Family Income	0.020	0.005	0.018	0.008	0.012	0.004	0.013	0.004
Intercept	-1.288	0.525	-0.931	0.874	-3.616	0.483	-2.863	0.646
Urban (age 17)	-0.179	0.097	0.508	0.161	0.153	0.088		
South (age 17)	-0.412	0.121	0.331	0.193	0.170	0.117		
West(age 17)	-0.398	0.125	0.209	0.211	-0.172	0.128		
Northeast (age 17)	0.213	0.124	0.136	0.222	0.400	0.107		
Local Unemployment (age 17)	2.887	1.756	5.344	2.872	3.948	1.628		
Local Long-run Unemployment	-10.992	4.347	-2.217	6.859	-5.127	4.051	-4.070	4.828
Age	0.043	0.019	-0.046	0.031	0.051	0.019	0.012	0.023
GED Passrate			-0.001	0.068				
Local 4-year College Tuition (age 17)					-0.261	0.059		
Local Unemployment (age 22)							-1.078	1.722
Local 4-year College Tuition (age 22)							-0.022	0.085
Urban (age 22)							0.058	0.131
South (age 22)							-0.107	0.153
West (age 22)							-0.469	0.180
Northeast (age 22)							0.031	0.154
Cognitive	0.823	0.093	1.011	0.138	0.846	0.092	0.833	0.114
Socioemotional	0.983	0.090	0.169	0.144	0.493	0.082	0.587	0.115
N	2242		522		1720		891	

Notes: The numbers in this table represent the estimated coefficients and standard errors associated with individual binary choice models of the sequential education model. Terminal schooling levels are highlighted in bold. Age in 1979 is included as a cohort control. We also included individual cohort dummies and it did not change the results. [a] Local long-run unemployment is the average local unemployment level over the previous 5 years. Local unemployment is the current unemployment rate. GED passrate is the estimated number of high school graduates able to pass the test on a single try given the state's passing standard. Unemployment variables, tuition, region dummies, and urban status are at age 17 for high school graduation, GED certification, and college enrollment choices. Tuition, unemployment variables, region dummies, and urban status are at age 22 for the choice to graduate from college.

To test the robustness of the measurement system and our identification strategy, several alternative specifications were estimated. For example, the measurement system was also estimated including models for early adverse behavior. Including these measures does not substantially change either the distribution of the factors or the loadings in the education and grade models. In Appendix Section I.2.4, we report that inclusion of a third factor, unrelated to cognitive and noncognitive measurements, but included in the schooling and outcome equations, does *not* substantially affect any of our empirical results.[48] This lends support to an empirical strategy of matching on mismeasured proxies for interpretable factors and correcting for measurement error.[49]

Table 4 presents the estimates of the schooling choice model. The cognitive factor loadings are statistically significant for all educational decisions. The socioemotional loadings are statistically significant in all decisions but the decision for GED certification. Figure 3 shows that the latent factors explain about 20%–40% of the variance in the educational models. Figure 4 shows the distribution of the factors by final schooling level. Figure 5 presents the probabilities of making the indicated educational choice at various levels of agent latent endowments. Individuals sort on both cognitive and socioemotional endowments into increasing schooling levels. The only exception are the GEDs, who have cognitive ability distributions similar to terminal high school graduates but socioemotional distributions similar to dropouts.

The latent endowments have statistically significant effects on labor market and health outcomes. Figure 6 plots the effects of the latent endowments on (log) wages, daily smoking, self-esteem, and physical health. The cognitive endowment affects all four outcomes, while the effect of the socioemotional endowment is statistically significant only in the equations for wages and smoking. Moving someone from the lowest decile to the highest decile in both cognitive and socioemotional ability, increases their wages by 0.6 log points, lowers the probability of being a smoker by 60%, increases their self-esteem by one standard deviation and increases their health by half a standard

[48]The third factor is assumed to be uncorrelated with the other two factors. Proof of identification of the model is straightforward using the analysis in Anderson and Rubin (1956).

[49]Appendix F shows the importance of accounting for measurement error in forming proxies for the factors.

deviation.

The estimates reveal clear evidence of sorting into education by both cognitive and socioemotional endowments. At the same time, these endowments have significant impacts on adult outcomes. Together these results imply strong selection biases in observed differences in outcomes by education level. This highlights the importance of accounting for observed and latent traits when estimating the causal impact of education.

6.2 The Effects of Educational Choices

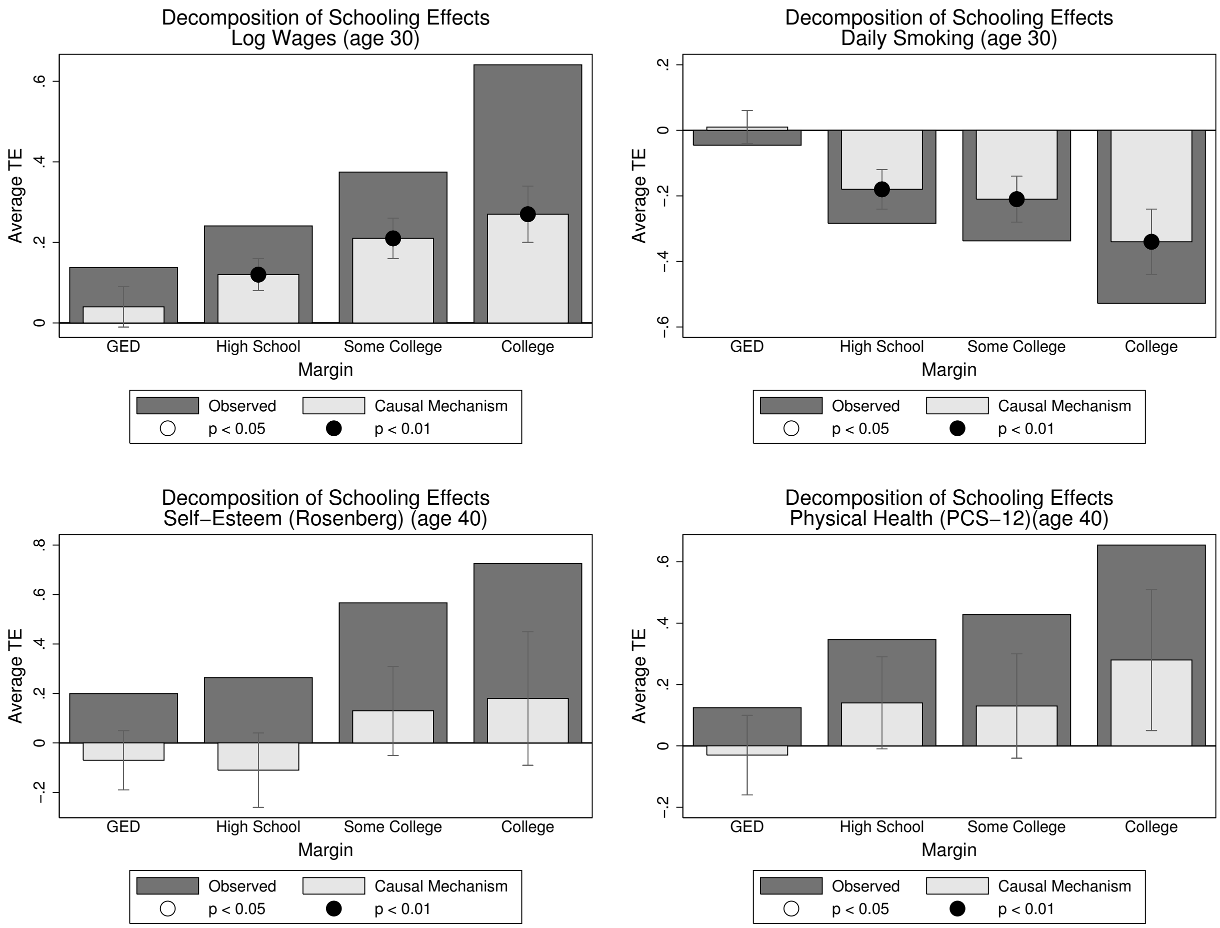
As discussed in Section 3, the causal effects of educational can be calculated by comparing outcomes between two final schooling levels or comparing final outcomes based on fixing decisions at a specified choice. We first compare the outcomes from a particular final schooling level s with those associated with being a high school dropout. The estimated treatment effects of education on log wages, smoking, physical health, and self-esteem are shown in Figure 7.[50] For each of the outcomes, the bars labeled "Observed" display the observed differences in the data. The bars labeled "Causal Mechanism" display the average treatment effect obtained from the comparison of the outcomes associated with a particular schooling level j relative to the high school dropout status.[51] This figure shows the observed earnings difference between dropouts and other educational groups. Since the "observed" bar only uses the individuals in two groups, the ATE is calculated over the same sub-population to increase comparability.

Our sequential model allows us to construct and analyze treatment effects by decision node. We also compute the gain to achieving (and possibly exceeding) the designated state inclusive of the continuation value and compare it to the outcome associated with not achieving the state. AMTE is the average treatment effect for the subpopulations

[50]These are calculated by simulating the mean outcomes for the designated state and comparing it with the mean-simulated outcome for the benchmark dropout state for the subpopulation of persons who are in either the designated state or the dropout state.

[51]Tables showing ATE for the full population; TT and TUT can be found in Section J of the Web Appendix.

Figure 7: Causal Versus Observed Differences by final schooling level



Notes: Each bar compares the outcomes from a particular schooling level j and the HS dropout status. The "Observed" bar displays the observed differences in the data. The "Causal Mechanism" bar displays the estimated average treatment effect (ATE) obtained from the comparison of the outcomes associated with a particular schooling level j relative to the HS dropout status. The ATE is calculated for those who have one of the final schooling levels being considered, so that the population is the same as the "Observed." The ATE for the full population and other margins is reported in the Web Appendix. The average treatment effect is $ATE_{s,0} \equiv \iint E(Y_s - Y_0|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{X}, \overline{\boldsymbol{\theta}} \,|\, H_0 + H_s = 1)$ where E is the expectation over idiosyncratic shocks to the outcomes Y_s and Y_0. The difference between the observed and causal treatment effect is attributed to the effect of selection and ability. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% levels is shown by open and filled circles on the plots, respectively.

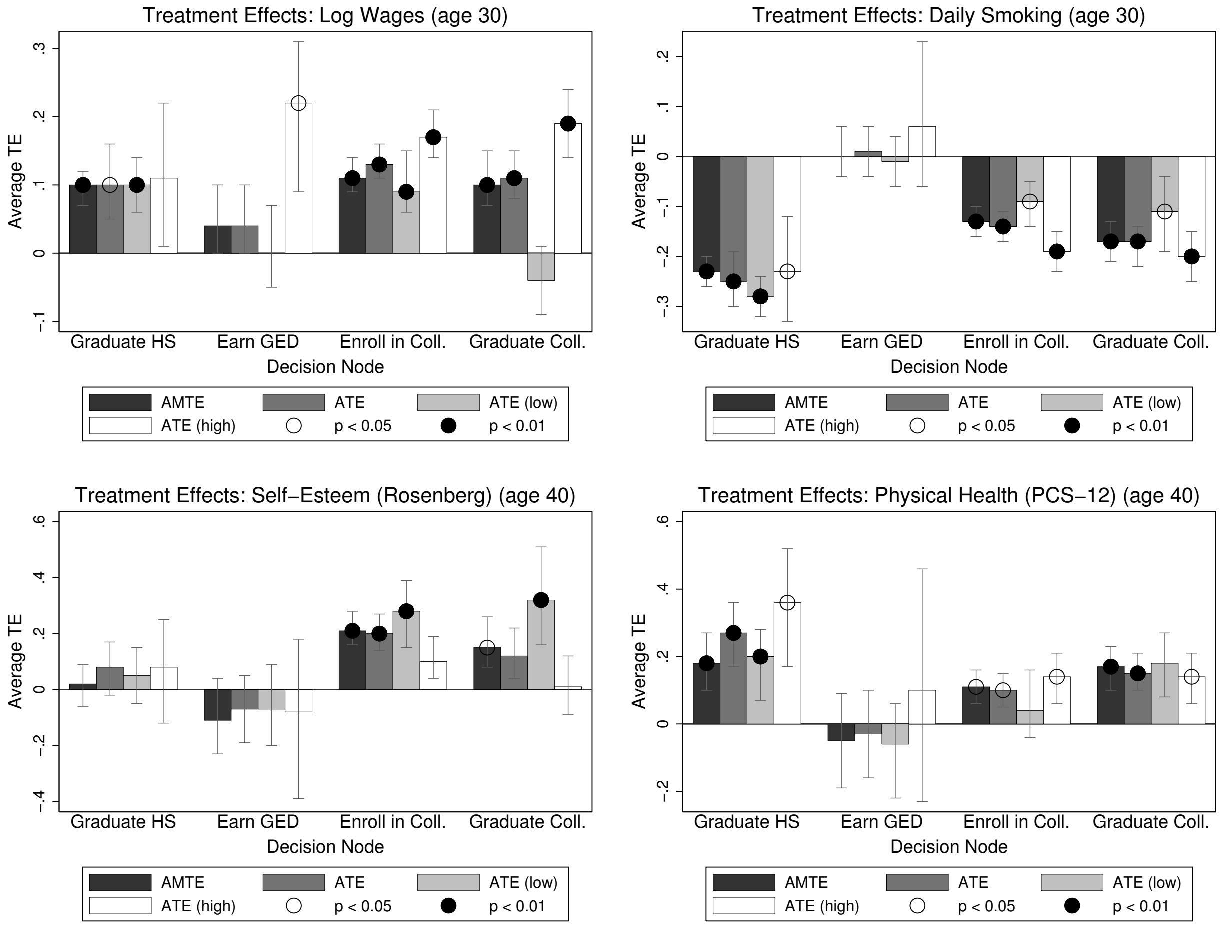
approximately indifferent to the two options of a particular choice.[52] The PRTE is defined as the average treatment effect for those induced to change final educational levels by a change in policy from p to p'. Again, the gains include continuation values.

The estimated treatment effects of educational decisions on log-wages, smoking, physical health and self-esteem are shown in Figure 8. Each figure presents the average effects of educational choices on the outcome of interest. The effects are presented as different bars in each figure, and they are defined as the differences in the outcome associated with the designated level and the one preceding it (not necessarily final or terminal schooling levels). Let $Y_{j'}$ denote the outcome (including continuation value) of choosing more schooling at a particular node. Let Y_j be the outcomes when no additional schooling is chosen at the particular node. Terminal schooling levels (GED certification and college graduation) do not provide any continuation value. At each node, ATE presents $E(Y_{j'} - Y_j | Q_{j,j'} = 1)$. ATE (high) and ATE (low) are the ATEs for different ability groups. The high- (low-) ability group is defined for individuals with both cognitive and socioemotional endowment above (below) the overall median of the full population. The table below the figure displays the fraction of individuals at each educational choice who are in the high- or low-ability group. The ATE is for the population who reaches the decision node. The ATE estimates for the entire population are reported in the Web Appendices J and K.

The panels in Figure 9 show that the estimated average treatment effect of getting a four-year degree depends on the latent ability of individuals for log wages, smoking, self-esteem and physical health. The figures display how treatment effects of graduating from college vary by decile of both cognitive and socioemotional endowments for each outcome. Unlike the previous graphs, these show average benefits by decile over the full population, rather than for the population that reaches each node. This makes deciles comparable across figures.

[52]We use the margin of indifference to be $|| \ I_{j,j'}/\sigma_j \ || \leq .02$.

Figure 8: Treatment Effects of Outcomes by Decision Node

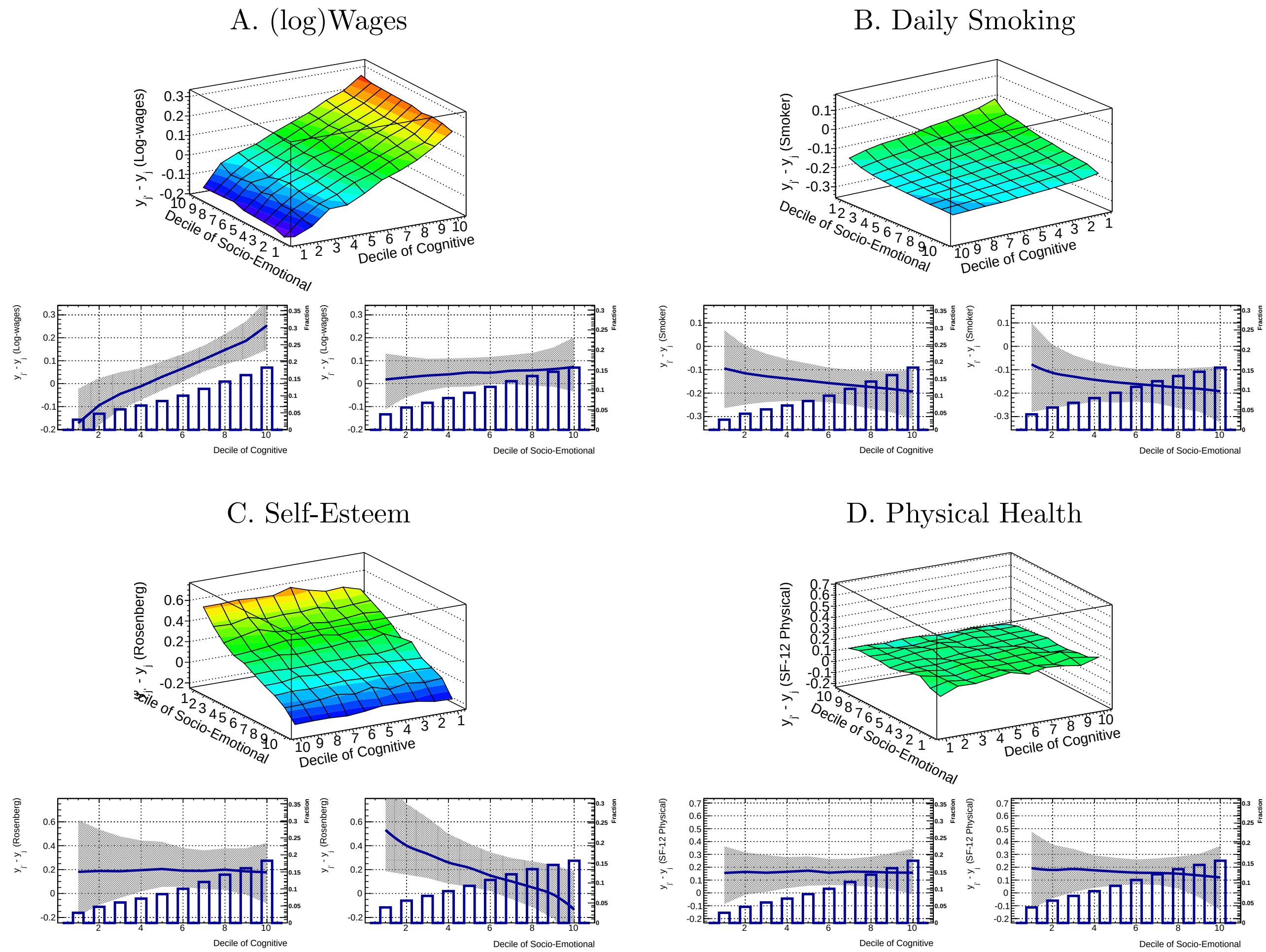


Notes: Let Y_j and $Y_{j'}$ denote the outcomes associated with schooling levels j and j', respectively. Importantly, each schooling level might provide the option to pursuing higher schooling levels. Only final schooling levels do not provide an option value. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% level are shown by hollow and black circles on the plots respectively. For two schooling levels j and j' associated with a particular educational choice, the ATE bar presents $E(Y_j - Y_{j'}|Q_{j,j'} = 1)$. The estimates for the full population are reported in the Web Appendix. AMTE presents the average $(Y_j - Y_{j'})$ for those who reach that decision node and are indifferent between j and j'. To be exact, the node-specific ATE is defined as, $ATE^D_{j,j'} \equiv \iint T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}]\, dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{X}, \overline{\boldsymbol{\theta}} \mid Q_{j,j'} = 1)$, where the individual's treatment effect is defined as $T_{j,j'}[Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}] \equiv E(Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ Q_{j,j'} = 1, Fix\ D_{j,j'} = 1) - E(Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ Q_{j,j'} = 1, Fix\ D_{j,j'} = 0)$ and the expectations of future outcomes are weighted by the probability of future educational choices: $E\left(Y|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{j,j'} = 1\right) \equiv \sum_s \Pr\left(s|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_{j,j'} = 1\right) \times E\left(Y_s|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}\right)$. The figure also presents the estimated ATE conditional upon endowment levels. The high- (low-) ability group is defined as those individuals with cognitive and socioemotional endowment above (below) the overall median. The fraction of individuals with low and high ability levels visiting each node are:

	Low Ability	High Ability
D_1: Dropping from HS vs. Graduating from HS	0.31	0.31
D_2: **HS Dropout** vs. **Getting a GED**	0.61	0.06
D_3: **HS Graduate** vs. College Enrollment	0.22	0.38
D_4: **Some College** vs. **4-year college degree**	0.14	0.51

In this table, final schooling levels are highlighted using bold letters.

Figure 9: Average Treatment Effect of Graduating from a Four-Year College by Outcome.



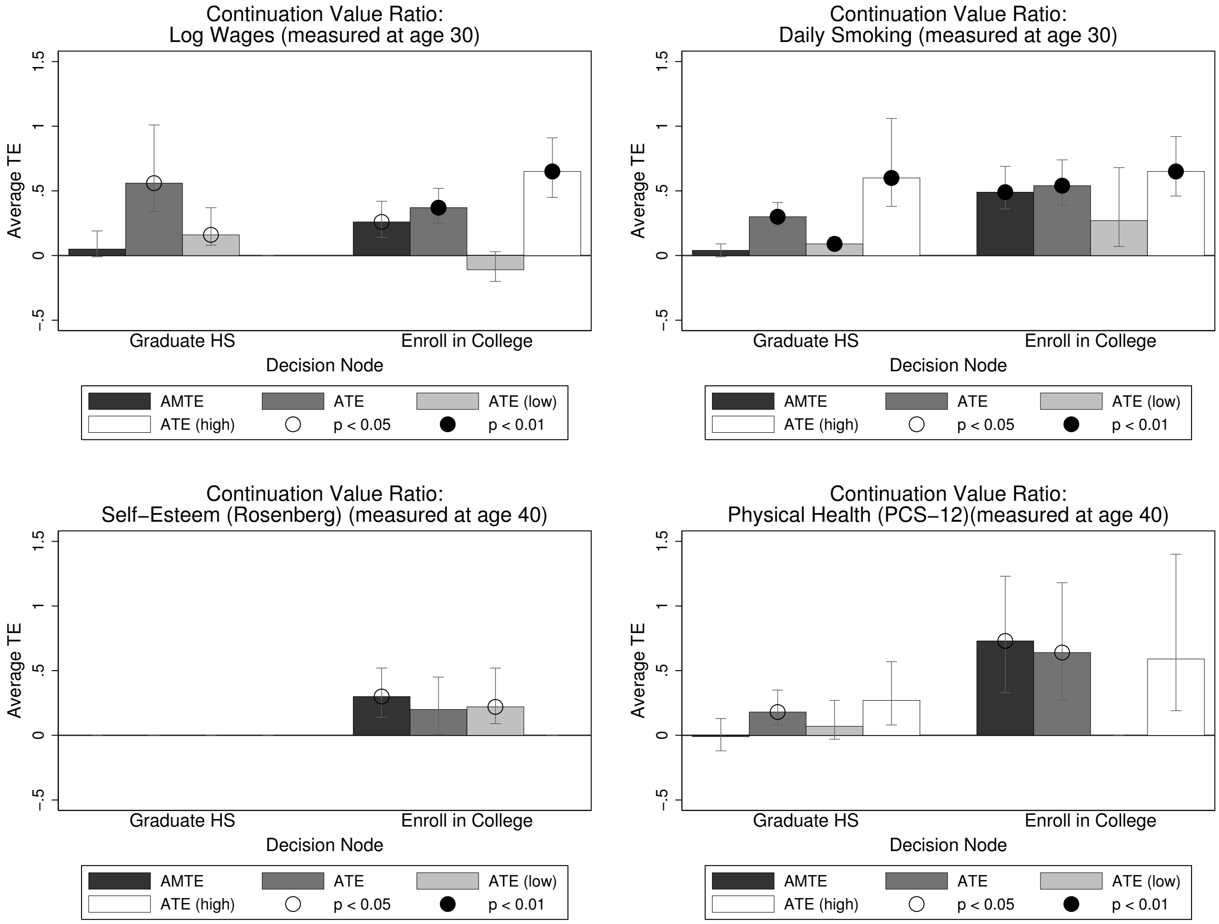
Notes: Each panel in this figure studies the average effects of graduating with a four-year college degree on the outcome of interest. The effect is defined as the differences in the outcome between those with a four-year college degree and those with some college. For each panel, let $Y_{somecoll}$ and $Y_{4-yr\ degree}$ denotes the outcomes associated with attaining some college and graduating with a four-year degree, respectively. For each outcome, the first figure (top) presents $E(Y_{4-yr\ degree} - Y_{somecoll}|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. The second figure (bottom left) presents $E(Y_{4-yr\ degree} - Y_{somecoll}|d^C)$ so that the socioemotional factor is integrated out. The bars in this figure display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving graduating from a four-year college. The last figure (bottom right) presents $E(Y_{4-yr\ degree} - Y_{somecoll}|d^{SE})$ and the fraction of individuals visiting the node leading to the educational decision involving graduating from a four-year college for a given decile of socioemotional endowment.

One benefit of schooling is access to further schooling.[53] Specifically, the choice to graduate from high school and the choice to enroll in college open up the doors for continued education. The continuation value of an educational choice is the probability of additional education times the benefits of that additional education. For high-ability individuals, the benefits of college may be large, and the probability of attending may be close to 1. For such individuals, the continuation value may constitute the bulk of the return to graduating from high school. For others, the probability or benefit of college may be much lower. Figure 10 reports the continuation value ratio—the proportion of the ATE accounted for by the continuation value. The continuation value ratio is reported both for high school graduation and enrollment in college. In Figure 10, the ratios are shown for high ability and low ability individuals, as well as the AMTE. The ratio is only shown if the ATE is statistically significant.[54]

[53]See Weisbrod (1962) and Comay, Melnik, and Pollatschek (1973).

[54]The continuation value ratio is only relevant for treatment effects that are significant. The ratio for imprecisely estimated numbers often have large, but insignificant values.

Figure 10: Treatment Effects: Ratio of Continuation Value to Treatment Effects



Notes: Estimates which have insignificant ATEs are not shown. (Taking ratios can lead to very large standard errors which compress the figures.) The ratio of average continuation value to average treatment effect is calculated by dividing the average continuation value by the average total treatment effect for those that reach the decision.. Continuation Value is the additional benefit gained through the option of pursuing additional education and is defined in Section 2. High-ability individuals are those in the top 50% of the distributions of both cognitive and socioemotional endowments. Low-ability individuals are those in the bottom 50% of the distributions of both cognitive and socioemotional endowments. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% level are shown by hollow and black circles on the plots respectively.

In general, the differences in outcomes between schooling levels are much larger when we do not control for observed variables and latent endowments. In most cases, the gains from education is increasing (in absolute value) with the schooling level, even after controlling for endowments. However, GED certification does not have significant causal effects on any of the outcomes.

Treatment Effects on Wages When comparing final education levels, the treatment effect for log wages is statistically significant for graduating from high school, some college achievement, and attaining a four-year college degree. The GED confers no benefits. About half of the observed difference in wages at age 30 are explained by observed variables and latent endowments.

When looking at the node-specific treatment effects on wages, higher educational attainment results in gains in wages, though low-endowment individuals gain very little from getting a four-year college degree. Figure 9 shows that it is individuals with high cognitive ability that capture most of the gains from a four-year degree. The GED confers no benefit except for individuals with large cognitive and socioemotional endowments – a group of individuals who rarely drop out of high school.

For log wages, the continuation value accounts for over half of the ATE from graduating from high school. While the total effect is relatively constant across treatment effects and endowment levels, low-endowment individuals benefit through the direct effect of being a high school graduate (see Figure 10). For the decision to enroll in college, less than half of the AMTE and ATE are continuation value, though more than half of the benefit for high-endowment individuals comes from continuation value.

Treatment Effects on Smoking Aside from the GED, education causally reduces smoking, even after accounting for unobserved heterogeneity. The latent endowments and observables account for one-third of the observed effect of education. Looking at the node-specific treatment effects, each educational decision also causally reduces smoking and it has an only weak dependence on latent endowments (see Figure 8 and Figure 9).

More than half of the average treatment effect of graduating high school and enrolling in college is derived from the continuation value for high-endowment individuals, while almost all of the treatment effect is coming from the direct effect for low-endowment individuals.

Treatment Effects on Self-Esteem When considering treatment effects by final schooling level, there is no statistically significant causal effect of schooling on self esteem. The estimates are more precise when looking at the dynamic treatment effects from specific educational decisions. Graduating from High School or obtaining the GED does not increase self-esteem, while enrolling in college and graduating with a four-year degree does. Interestingly, the gain is heterogeneous by ability, with the low-ability individuals getting larger gains. Figure 9 shows it is individuals with low socioemotional ability that capture most of the gains from a four-year degree. The improvements in self-esteem and self-mastery from enrolling in college are explained almost completely by the direct effect of having attended college, though a small and statistically significant portion of the gains are from continuation value for the low-endowment ATE and the AMTE (see Figure 10).

Treatment Effects on Physical Health The estimated effects by final schooling level are imprecise. There is no statistically significant effect from final education level on physical health. On the other hand, the estimated causal effects of most educational decisions are statistically significant. Graduating from High School and graduating with a four-year degree both show statistically significant improvements in physical health. There is no evidence of heterogeneous treatment effects by ability. Finally, the continuation value accounts for a portion of the average physical health benefits from graduating from high school and enrolling in college.

6.3 Policy Relevant Treatment Effects

While estimating the returns to particular educational choices is informative, it does not correspond to any particular policy. The PRTE allows us to identify who would be induced to change educational choices under a particular policy change, and how these individuals would benefit on average. As an example, we simulate the response to a policy intervention that provides a one standard deviation subsidy towards early college tuition (approximately $\$7,500$ dollars per year). Column 1 of Table 5 provides the average treatment effect for those that are induced to change education levels by the tuition subsidy. Since tuition at age 17 only enters the choice to enroll in college, the subsidy will only induce high school graduates to change their college enrollment decisions. Those induced to enroll may then go on to choose to graduate with a four year degree.[55] Columns 2 and 3 of Table 5 decompose the PRTE into the average gains for those induced to enroll that then go on to earn 4 year degrees and the average gains for those that do not. Except for self-esteem, the PRTE is larger for those that then go on to earn 4 year degrees.

Figure 11 shows which individuals are induced to enroll in college within the deciles of the distribution of $\epsilon_{1,3} = \theta\alpha_{1,3} - \nu_{1,3}$, conditional on $Q_{1,3} = 1$, the unobserved components of heterogeneity acted upon by the agent but unobserved by the econometrician. The policy induces some individuals in every decile to switch, but places more weight on those in the middle deciles of the distribution. The figure further decomposes the effect of those induced to switch into the effect for those who go on to graduate with four year degrees and the effect for those who do not. We see that those induced to switch in the top deciles are more likely to go on to graduate.[56]

Table 5: PRTE: Standard Deviation Decrease in Tuition

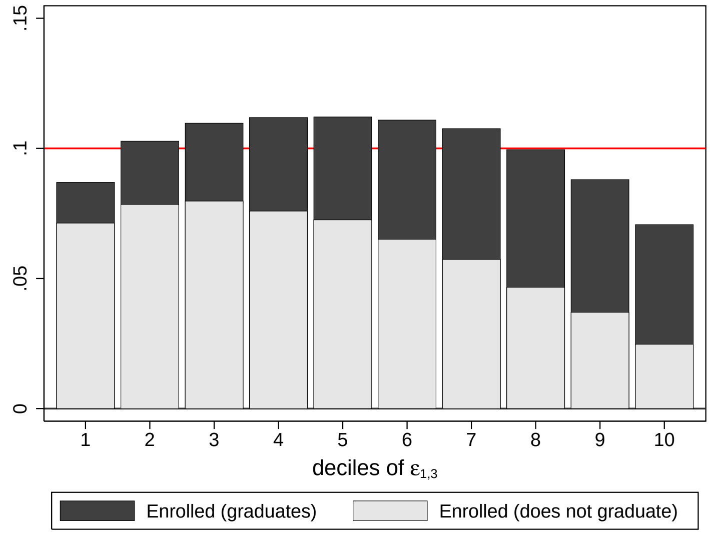
	PRTE	4-year degree	no 4-year degree
Log Wages	0.12	0.14	0.11
PV Log Wages	0.13	0.13	0.12
Physical Health	0.10	0.13	0.09
Self-Esteem	0.22	0.15	0.27
Smoking	-0.13	-0.16	-0.11

Notes: Table shows the policy relevant treatment effect (PRTE) of reducing **early** tuition by a standard deviation (approx. \$7,500). The PRTE is the average treatment effect of those induced to change educational choices as a result of the policy: $PRTE_{p,p'} \equiv \iint E(Y^p - Y^{p'}|\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}), dF_{\boldsymbol{X},\boldsymbol{\theta}}(\boldsymbol{X}, \overline{\boldsymbol{\theta}}|S(p) \neq S(p')\}$. Column 1 shows the overall PRTE. Column 2 shows the PRTE for those induced to enroll by the policy who then go on to complete 4-year college degrees. Column 3 shows the PRTE for individuals induced to enroll but who do not complete 4-year degrees.

[55] Models were estimated that include tuition as a determinant of the high school graduation decision. However, estimated effects of tuition on high school graduation are small and statistically insignificant.

[56] Note that there is little evidence of forward-looking behavior in terms of tuition.

Figure 11: PRTE: Who is induced to switch



Notes: The figure plots the proportion of individuals induced to switch from the policy that lay in each decile of $\epsilon_{1,3}$, where $\epsilon_{1,3} = \theta\alpha_{1,3} - \nu_{1,3}$. $\epsilon_{1,3}$ is the unobserved component of the educational choice model. The deciles are conditional on $Q_{1,3} = 1$, so $\epsilon_{1,3}$ for individuals who reach the college enrollment decision. The bars are further decomposed into those that are induced to switch that then go on to earn 4-year degrees and those that are induced to switch but do not go on to graduate.

The $7,500 subsidy induces 13% of high school graduates who previously did not enroll to enroll in college. Of those induced to enroll, more than a third go on to graduate with a 4 year degree. For outcomes such as smoking, the benefits are larger for those who graduate with a 4-year degree, while for outcomes such as self-esteem, those that enroll, but choose not to earn 4 year degrees have larger benefits.

Using the estimated benefits, we can calculate if the policy is cost effective. As a limitation of our model, we can only estimate the monetary costs and cannot estimate psychic costs. We can determine if the monetary gains in the present value of wages at age 18 is greater than the $7,500 subsidy. Given a PRTE of 0.13 for log present value of wage income, the average gains for those induced to enroll was $11,275.[57] If the subsidy was given for the first two years of college, then the policy is not cost-efficient on average. If the subsidy must also be offered to those already enrolled, then monetary costs greatly outweigh the estimated earnings benefits, never mind any psychic costs.

[57] Wages are divided by $10,000 prior to taking logs

While the intervention does not pay for itself through increased income, there are benefits for several other life-outcomes such as reduced smoking, improved self-esteem, and improved physical health.

6.4 The Channels of Influence of Cognitive and Socioemotional Skills

The traditional literature focuses on cognitive ability as a major determinant of schooling and the outcomes of schooling. Much less is known about the importance of socioemotional factors, although a growing literature establishes their predictive power. We present new evidence on the channels through which cognitive and socioemotional skills operate. As shown in Figure 5 and Table 4, both factors are important in explaining schooling choices at all decision nodes with the exception of socioemotional skills in explaining the decision of dropouts to obtain GEDs.

Controlling for selection into schooling levels, we can determine if, after fixing schooling, there are additional effects of cognitive and socioemotional skills on outcomes. Is the main effect of these skills through educational attainment or are there additional effects of the traits on outcomes beyond their effects on schooling?

The models for outcomes are estimated twice: (a) excluding any effect of the latent factors on outcomes except through their effects on schooling and (b) allowing the latent factors to have additional effects beyond their effects on schooling. A comparison of the estimates is informative.

Table 6 reports, for each set of outcomes, tests of whether the loadings on the latent factors at all schooling levels (HS Dropout—Coll. Graduate) are all zero (column a) or whether they are equal to their levels in the unconditional model (column b). For most outcomes, cognition has a statistically significant effect beyond its educational attainment. For physical health and self-esteem, the effects of socioemotional skills are not precisely estimated. We cannot reject the hypothesis that across all s, the socioemotional loadings are jointly equal to 0. We also cannot reject that they are the

Table 6: Estimated Factor Loadings on Cognitive and Socioemotional Factors by Outcome and Schooling Level

Variables	Tests	
	p-val$^{(a)}$	p-val$^{(b)}$
Log Wages		
Cognitive	0.000	0.001
Socioemotional	0.221	0.006
Smoking (Age 30)		
Cognitive	0.133	0.002
Socioemotional	0.037	0.007
Physical Health (Age 40)		
Cognitive	0.155	0.047
Socioemotional	0.513	0.275
Self-Esteem		
Cognitive	0.000	0.004
Socioemotional	0.465	0.469

Notes: (a) shows p-values from a likelihood ratio test against the null hypothesis that the factor loadings for the conditional models are jointly equal to zero. (b) shows the p-value from a likelihood ratio test against the null hypothesis that the factor loadings for the conditional models are jointly equal to the factor loading of the unconditional model. A table reporting the factor loadings for the unconditional and conditional models and the results for other outcomes can be found in Section C.5 in the Web Appendix.

same as the loadings for the factors on outcomes that do not control for s. However, in the case of smoking, it appears that the socioemotional trait has a statistically significant effect beyond its effect on schooling. Finally, in terms of wages, socioemotional trait operates primarily through its affect on schooling.

7 Summary and Conclusions

This paper formulates and estimates a sequential model of educational choices and their consequences for wages, health and healthy behaviors, allowing for heterogeneity among agents in both observed and unobserved characteristics. We estimate the causal impact of education on health and labor market outcomes when responses to treatment vary among observationally identical persons who select into schooling levels on the basis of their idiosyncratic responses. We proxy the unobservables producing dependence between choices and outcomes using a variety of measurements. We adjust for the measurement error arising from using proxies. We use multiple sources of identification to secure our estimates. Each educational choice has exclusion restrictions that affect choices but not the outcomes produced from the choice. These can be used as instruments to identify the model. They allow us to control for unobservables that generate dependence between choices and outcomes that are not proxied by our measurements.

Unlike most of the literature on the treatment effects of education, we analyze a model with multiple schooling choices that recognizes the fundamentally nonlinear effect of schooling on a variety of outcomes. Our paper defines and estimates a variety of novel treatment effects, including treatment effects that account for the continuation values associated with future educational choices which we find to be empirically important for most transitions.

Our empirical results show that there is strong sorting into schooling levels on both cognitive and socioemotional endowments. Overall, we find that wages and the health and healthy behaviors of persons is enhanced by high school graduation, college enrollment, and graduation from a 4-year college. We find that the causal effects of schooling differ by ability level. For example, when considering the benefits of a four-year degree, only high-ability individuals gain higher wages, while only low-ability individuals gain higher self-esteem. In general, observed differences by educational attainment diminish when we control for observables and latent abilities.

The model estimated in this paper is a halfway house between the treatment effect literature and the structural literature on dynamic discrete choice. By modeling the latent variable structure, we improve on LATE by identifying the groups affected by variations in instruments and policies. We find strong evidence that agents select into schooling based on their idiosyncratic responses to schooling. Our model requires no assumptions about how agents form expectations or what they seek to maximize that are routinely invoked in the structural economics literature.[58]

By estimating a sequential model of schooling in a single framework, we are able analyze the *ex post* returns to education for people at different margins of choice and to analyze a variety of interesting policy counterfactuals. We are able to characterize who benefits from education across a variety of market and nonmarket outcomes. We decompose these benefits into direct components and indirect components arising from continuation values that we estimate to be substantial. Standard estimates of the benefits of education based on direct components underestimate the full benefit of education.

[58]This generality comes at a cost. Agent information sets are not precisely specified or identified and we cannot estimate psychic costs or net returns.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2013): "Education Policy and Intergenerational Transfers in Equilibrium," Working Paper 18782, National Bureau of Economic Research.

Abbring, J. H., and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 5145–5303. Elsevier, Amsterdam.

Adams, S. J. (2002): "Educational Attainment and Health: Evidence from a Sample of Older Adults," *Education Economics*, 10(1), 97–109.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Wößmann, vol. 4. Elsevier, Amsterdam, Forthcoming.

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11(1), 48–83.

Anderson, T., and H. Rubin (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5*, ed. by J. Neyman, pp. 111–150. University of California Press, Berkeley.

Angrist, J. D., and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

Arcidiacono, P., and P. B. Ellickson (2011): "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, 3, 363–394.

ARENDT, J. N. (2005): "Does education cause better health? A panel data analysis using school reforms for identification," *Economics of Education Review*, 24(2), 149–160.

ARKES, J. (2003): "Does Schooling Improve Adult Health?," Unrestricted Draft DRU-3051, RAND, Santa Monica, CA.

AULD, M. C., AND N. SIDHU (2005): "Schooling, Cognitive Ability and Health," *Health Economics*, 14(10), 1019–1034.

BECKER, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. National Bureau of Economic Research, distributed by Columbia University Press, New York.

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND B. TER WEEL (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4), 972–1059.

BORGHANS, L., B. H. H. GOLSTEYN, J. J. HECKMAN, AND J. E. HUMPHRIES (2011): "Identification Problems in Personality Psychology," *Personality and Individual Differences*, 51(Special Issue on Personality and Economics), 315–320, E. Ferguson, J.J. Heckman, and P. Corr, editors.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American statistical association*, 90(430), 443–450.

CAMERON, S. V., AND J. J. HECKMAN (1987): "Son of CTM: The DCPA Approach Based on Discrete and Factor Structure Models," Unpublished working paper, University of Chicago, Department of Economics.

——— (1993): "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11(1, Part 1), 1–47.

——— (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–99.

Card, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 5, pp. 1801–1863. North-Holland, New York.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

Carneiro, P., K. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2010): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica*, 78(1), 377–394.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–2781.

Cawley, J., and C. J. Ruhm (2012): "The Economics of Risky Health Behaviors," *Handbook of Health Economics*, 2, 95–199.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B. Elsevier, Amsterdam.

Comay, Y., A. Melnik, and M. A. Pollatschek (1973): "The Option Value of Education and the Optimal Path for Investment in Human Capital," *International Economic Review*, 14(2), 421–435.

Conti, G., J. J. Heckman, P. Pinger, and A. Zanolini (2009): "Intergenerational Transmission of Inequality: Maternal Skills, Prenatal Risk Behavior and Child Birth Weight," Unpublished manuscript, University of Chicago, Department of Economics.

Conti, G., J. J. Heckman, and S. Urzua (2010): "The Education-Health Gradient," *American Economic Review: Papers and Proceedings*, 100(2), 1–5.

Cunha, F., and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4), 738–782.

Cunha, F., J. J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings, The 2004 Hicks Lecture," *Oxford Economic Papers*, 57(2), 191–261.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

Cutler, D. M., and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29(1), 1–28.

Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2010): "What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades," Under review, *Journal of Educational Psychology*.

Duckworth, A. L., and M. E. P. Seligman (2005): "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents," *Psychological Science*, 16(12), 939–944.

Eckstein, Z., and K. I. Wolpin (1989): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *Journal of Human Resources*, 24(4), 562–598.

Eisenhauer, P., J. J. Heckman, and S. Mosso (2013): "Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments," Under review, *International Economic Review*.

Eisenhauer, P., J. J. Heckman, and E. J. Vytlacil (2014): "Generalized Roy Model and Cost-Benefit Analysis of Social Programs," Unpublished manuscript, University of Chicago, Department of Economics.

Gandek, B., J. E. Ware, N. K. Aaronson, G. Apolone, J. B. Bjorner, J. E. Brazier, M. Bullinger, S. Kaasa, A. Leplege, L. Prieto, and M. Sullivan (1998): "Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results from the IQOLA Project," *Journal of Clinical Epidemiology*, 51(11), 1171–1178.

Geweke, J., and M. Keane (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3463–3568. North-Holland, New York.

Grossman, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223–255.

——— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, vol. 1, pp. 347–408. Elsevier, Amsterdam.

——— (2006): "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 10, pp. 577–633. Elsevier, Amsterdam.

——— (2008): "The Relationship between Health and Schooling: Presidential Address," *Eastern Economic Journal*, 34(3), 281–292.

Grossman, M., and R. Kaestner (1997): "Effects of Education on Health," in *The Social Benefits of Education*, ed. by J. R. Behrman, and N. Stacey, pp. 69–124. University of Michigan Press, Ann Arbor, MI.

GULLONE, E., AND S. MOORE (2000): "Adolescent risk-taking and the five-factor model of personality," *Journal of Adolescence*, 23(4), 393–407.

HAAVELMO, T. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11(1), 1–12.

HANSEN, K. T., J. J. HECKMAN, AND K. J. MULLEN (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1-2), 39–98.

HANSEN, L. P., AND T. J. SARGENT (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2(1), 7–46.

HECKMAN, J., AND R. PINTO (2013): "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs," Forthcoming, *Econometric Reviews*.

HECKMAN, J. J. (1981): "The Empirical Content of Alternative Models of Labor Earnings," Unpublished manuscript, University of Chicago, Department of Economics.

——— (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32(3), 441–462, Addendum published vol. 33 no. 1 (Winter 1998).

——— (2008): "The Principles Underlying Evaluation Estimators with an Application to Matching," *Annales d'Economie et de Statistiques*, 91-92, 9–74.

HECKMAN, J. J., J. E. HUMPHRIES, AND T. KAUTZ (2014): "The Economic and Social Benefits of GED Certification," in *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, ed. by J. J. Heckman, J. E. Humphries, and T. Kautz. University of Chicago Press, Chicago.

HECKMAN, J. J., AND T. KAUTZ (2012): "Hard Evidence on Soft Skills," *Labour Economics*, 19(4), 451–464, Adam Smith Lecture.

——— (2014): "Fostering and Measuring Skills: Interventions That Improve Character and Cognition," in *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press, Chicago, IL, Forthcoming.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings Equations and Rates of Return: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, and F. Welch, chap. 7, pp. 307–458. Elsevier, Amsterdam.

Heckman, J. J., and S. Mosso (2014): "The Economics of Human Development and Social Mobility," Forthcoming, *Annual Reviews of Economics*.

Heckman, J. J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136(2), 341–396.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 1–35.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Heckman, J. J., and S. Urzua (2010): "Comparing IV With Structural Models: What Simple IV Can and Cannot Identify," *Journal of Econometrics*, 156(1), 27–37.

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88(3), 389–432.

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96(8), 4730–4734.

——— (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

——— (2007a): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 4779–4874. Elsevier, Amsterdam.

——— (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 4875–5144. Elsevier, Amsterdam.

Herrnstein, R. J., and C. A. Murray (1994): *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press, New York.

Imbens, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48(2), 399–423.

Keane, M. P., P. E. Todd, and K. I. Wolpin (2011): *The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications*, vol. 4 of *Handbook of Labor Economics*. Elsevier.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

Lleras-Muney, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72(1), 189–221.

Lochner, L. (2011): "Nonproduction Benefits of Education: Crime, Health, and Good Citizenship," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, vol. 4, pp. 183–282. Amsterdam: North-Holland.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302.

——— (1974): *Schooling, Experience and Earnings*. Columbia University Press for National Bureau of Economic Research, New York.

OREOPOULOS, P., AND K. G. SALVANES (2011): "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives*, 25(1), 159–184.

QUANDT, R. E. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53(284), 873–880.

SILLES, M. A. (2009): "The causal effect of education on health: Evidence from the United Kingdom," *Economics of Education Review*, 28(1), 122–128.

SPASOJEVIC, J. (2003): "Effects of Education on Adult Health in Sweden: Results from a Natural Experiment," Unpublished manuscript, City University of New York Graduate Center.

WARE, JOHN E., J., M. KOSINSKI, AND S. D. KELLER (1996): "A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity," *Medical Care*, 34(3), 220–233.

WEISBROD, B. A. (1962): "Education and Investment in Human Capital," *Journal of Political Economy*, 70(5, Part 2: Investment in Human Beings), 106–123.

WILLIAMS, B. (2011): "Identification of Factor Models," Unpublished manuscript, University of Chicago, Department of Economics.