Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking: Web Appendix*

James J. Heckman
University of Chicago
and the American Bar Foundation

John Eric Humphries
University of Chicago

Gregory Veramendi
Arizona State University

March 1, 2016

*James J. Heckman: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-702-0634; fax: 773-702-8490; email: jjh@uchicago.edu. John Eric Humphries: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-980-6575; email: johneric@uchicago.edu. Gregory Veramendi: Arizona State University, 501 East Orange Street, CPCOM 412A, Tempe, AZ 85287-9801; phone: 480-965-0894; email: gregory.veramendi@asu.edu. This paper was presented at the Becker Friedman Institute conference in honor of Gary Becker, October 30, 2014. It was also presented as the Sandmo Lecture at the Norwegian School of Economics, January 13, 2015. We thank Chris Taber for insightful comments on an early draft. We also thank Ariel Pakes and other participants at a Harvard Labor Economics Workshop in April, 2014, for helpful comments on a previous draft. We thank Eleanor Dillon and Matthew Wiswall for comments received at a seminar at Arizona State University, February, 2015. We thank the special editor, Ed Lazear, and an anonymous referee for helpful comments. We also thank Jessica Yu Kyung Koh, Joshua Shea, Jennifer Pachon, and Anna Ziff for 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; NIH grants NICHD R37HD065072, NICHD R01HD054702, and NIA R24AG048081; an anonymous funder; Successful Pathways from School to Work, an initiative of the University of Chicago's Committee on Education funded by the Hymen Milgrom Supporting Organization; and the Human Capital and Economic Opportunity Global Working Group, an initiative of the Center for the Economics of Human Development, affiliated with the Becker Friedman Institute for Research in Economics, and funded by the Institute for New Economic Thinking. Humphries acknowledges the support of a National Science Foundation Graduate Research Fellowship. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The Web Appendix for this paper is https://heckman.uchicago.edu/eff-ed-earn-health.

Contents

List of Figures

List of Tables

A Web Appendix

A.1 A Summary of the Related IV Literature

This section provides a brief review of the instrumental variable literature on the returns to education for the outcomes considered in our paper.

Log Wages Much of the literature that focuses on the returns to log wages from schooling estimates Mincer model (1) and instruments for years of completed schooling. The LATE estimator is a weighted average of the returns to those induced to make changes in their final education across different schooling transitions divided by the change in years of schooling. It does not recover the causal effects of agents taking specific transitions. Typical applications of LATE do not report the Heckman-Vytlacil (2007) IV weights that identify the populations affected.[1] This makes direct comparisons of results from the IV literature with our results from our analysis difficult. Card (1999) reviews a number of papers that use changes in school policy (such as changes in mandatory schooling). These papers estimate LATEs of a version of model (1) ranging from 0.06 to 0.153, although most are around 0.10. We are not aware of any IV studies with multiple margins where the returns at the various margins are identified.

Smoking Grimard and Parent (2007) and de Walque (2007) use the Vietnam draft lottery number as an instrument for the effect of college enrollment on smoking. Both papers estimate local average treatment effects and find that an additional year of schooling lowers the probability of smoking by 8% and 4–9%, respectively. In other work, Koning et al. (2015) use twin studies to estimate the effect of education on smoking, and Sander (1998) uses matching on observables. The first paper finds a reduction of 4% a year, while the second finds that college enrollment reduces the probability of smoking by over 50%.

[1] See Section 6.3 in the main text.

Health Limits Work The literature is mixed on the causal relationship between education and work limitations due to health. Auld and Sidhu (2005) find a causal link between education and health-related work limitations, but only for the low-ability populations and those with low levels of education. Arendt (2005) uses state variation in teenage unemployment rates and finds a LATE of 2.6% for education on health-related work limitations. Other papers such as Adams (2002) and Oreopoulos (2006) look at the closely related outcome of "health limits activity" and find mixed or positive impacts of education.[2]

A.2 Description of the Data Used

A.2.1 Data

We estimate our model on the 1979 National Longitudinal Survey of Youth (NLSY79). These data are widely used to estimate the returns to schooling.[3] 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 has a variety of adult outcome measures including income and health. This sample 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.[4] We report results for samples that pool race groups, but enter dummy variables to control for mean differences.

[2] See Albouy and Lequien (2009) for an overview of these papers.

[3] See, e.g., Keane and Wolpin (1997, 2000); Keane (2002); Keane and Wolpin (2001); Heckman and Raut (2013); Sullivan (2010); Gladden and Taber (2000); Kane and Rouse (1995); Kling (2001); Belzil and Hansen (2002).

[4] Respondents were dropped from the analysis if they did not have valid ASVAB scores, missed multiple rounds of interviews, 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. Value in previous years were used when not available for a given year (such as region of residence). Responses from adjacent years were used for some outcomes when outcome variables were missing. Mother's education and father's education were imputed when missing. See Web Appendix Section A.2 for the analysis of the deleted observations.

We estimate models with the four different transitions and five final schooling levels discussed in Section A.2.1 (see Figure 2). Education at age 30 is treated as the respondent's final schooling level.[5]

A.2.2 Outcomes

This paper considers the effect of education on log wages, log present value of wage income, smoking, and health limits work.[6]

As a measure of adult physical health, we follow Auld and Sidhu (2005) and use an indicator of whether health limits the type or amount of work individuals are able to perform.[7] We study smoking at age 30 as an additional measure of healthy behavior. It is a self-reported, binary variable recording whether the individual smoked daily at age 30.

We analyze the effect of education on log wages at age 30 as a traditional benchmark. We also estimate the causal effect of education on the log present value (PV) of wage income, as it is closely related to the outcome considered in many structural models. Present value of wage income is defined as the present value (at age 18) of earnings from ages 18 to 40, discounted at 5%.

A.2.3 Schooling Levels

We consider four different transitions and five final schooling levels. The transitions studied are (i) enrolled in high school, deciding between graduating from high school and dropping out from high school; (ii) high school dropouts deciding whether or not to get the GED; (iii) high school graduates deciding whether or not to enroll in college; and (iv) college students deciding whether or not to earn a four-year degree. Consequently, the final schooling levels are (I) high school dropout; (II) GED; (III) high school graduate; (IV) some college; and (V)

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

[6] The literature focuses primary attention on the effect of mortality and on smoking. See the summary of this literature provided in Web Appendix A.1.

[7] We construct an indicator variable denoting whether or not individuals report that their health limits the type or amount of work they can perform between 1994 (when respondents are on average 34 years old) and 2010. This variable avoids the potential subjectivity of self-reported health measures.

four-year college degree. Education at age 30 is treated as respondent's final schooling level.[8]

A.2.4 Measurement System

The cognitive and socio-emotional factors in the model are identified from the joint estimation of the educational choices of agents as well as a supplemental measurement system of tests and other early-life outcomes. Sub-tests from the Armed Services Vocational Aptitude Battery (ASVAB) are used as measures of cognitive ability. Specifically, we consider the scores from Arithmetic Reasoning, Coding Speed, Paragraph Comprehension, Word Knowledge, Math Knowledge, and Numerical Operations.[9]

We also use participation in minor risky or reckless activity in 1979 in the measurement system for both the socio-emotional endowment and cognitive endowment. Estimated results change little when only allowing risky and reckless activity to depend on the socio-emotional factor.[10]

Many psychologists use a socio-emotional taxonomy called the Big Five (John et al., 2008). This is an organizing framework that categorizes personality traits into five categories. The five traits are extraversion, agreeableness, conscientiousness, neuroticism, and openness. A growing body of work suggests that these traits and other socio-emotional traits play key roles in academic success. Borghans et al. (2011) and Almlund et al. (2011) show that 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. Duckworth et al. (2012) report three studies to show that self-control predicts grades earned in middle school better than IQ across racial and socio-economic groups. Farsides

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

[9] A subset of these tests are 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 et al. (2011).

[10] This is a binary variable, which is one if an agent answered 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?"

and Woodfield (2003), Conard (2006), and Noftle and Robins (2007) find that Big Five traits positively predict grades and academic success. These studies find predictive power after controlling for previous grades or test scores. In these studies, the benefits of personality traits are mediated through behaviors such as increased attendance or increased academic effort. A meta-analysis by Credé and Kuncel (2008) finds that study habits, endowments, and attitudes have similar predictive power as standardized tests and previous grades in predicting college performance.

Academic success (such as GPA) depends on cognitive ability, but also depends strongly on socio-emotional traits such as conscientiousness, self-control, self-discipline, and motivation. We allow 9th grade GPA to depend on both cognitive and socio-emotional factors. Much of the variance not explained through test scores has been shown to be related to socio-emotional traits. Socio-emotional endowments are measured in part by their contribution towards 9th grade GPA in reading, social studies, science, and math.

GPA by grade and subject is constructed from high school transcript records. Up to 64 courses were recorded from school transcripts and included year taken, grade level taken, a class identification code, and the grade received. Using the class identification code, we identified all courses taken in either reading, social studies, science, or math in 9th grade and constructed subject level GPAs.

As a robustness check for our measure of socio-economic endowments, we include five additional measures of adverse adolescent behavior to check our interpretation of the non-cognitive factor.[11] 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 sexual intercourse before age 15. For violent behavior, we control for the potential effect of schooling. We estimate the cognitive and socio-emotional distributions jointly with the educational choice system to account for the effect of schooling

[11] Gullone and Moore (2000) present a line of research which studies the relationship between personality traits and adolescent risk behavior. Our five additional measures of early adverse behavior help demonstrate that our socio-emotional factor is capturing traits that then explain these observed behaviors in an expected manner.

at the time of the measurement on measures of ability following the procedure developed in Hansen et al. (2004).

A.2.5 Control Variables

The variables used to control for observed characteristics depend on the timing and nature of the decision being made. In 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 determined.[12] For log wages at age 30, we additionally control for local economic conditions at age 30. When region of residence or urban status are not available for the age of a particular measure or outcome, the answer from previous or following surveys are used.

The educational choice models include additional choice-specific covariates. Following Carneiro et al. (2011), we control for both long-run economic conditions and contemporaneous deviations from those conditions. Controlling for the long-run local economic environment, local unemployment deviations capture contemporaneous economic shocks. The model for the choice to GED certify additionally controls for the difficulty of getting the GED within the state of residence in 1988.[13] The choices to enroll in college and graduate from college control for local four-year college tuition at ages 17 and 22, respectively. When an instrument is missing for a particular age, the value from the previous or proceeding year is used.

The equation system for GPA controls for the variables used in all of our analyses, except for region dummies which are not available prior to 1979. The GPA model alternatively controls for urban status at age 14 and Southern residence at age 14. The ASVAB test scores models control for the standard controls, age, and age squared. As previously noted above, the ASVAB tests are estimated separately by education at the time of the test. Risky

[12]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.

[13]GED difficulty is proxied by the percent of high school graduates estimated to be able to pass the test in one try given the state's chosen average and minimum score requirements.

behavior in the 1979 model controls for the standard controls, age and age squared. The risky behavior measure is also estimated by educational group, but due to data limitations, pools high school graduates and those enrolled in college in 1979.

The equations for log wages at age 30 control for race, parents' education, parents' income, broken home status, number of siblings, region of residence at age 30, urban environment at age 30, dummies for living in the South or in an urban area at age 14, and local unemployment rates at age 30. Smoking at age 30 controls for the same variables, but excludes the local unemployment rate.

Present value of wages and health limits work control for race, parents' education, parents' income, broken home status, number of siblings, and 1979 region of residence and urban environment dummies.

A.2.6 Constructing the Data

As a baseline, our National Longitudinal Survey of Youth 1979 dataset uses the NLSY79 dataset used in Heckman et al. (2006), and Heckman (2001). Furthermore, we use instruments from Carneiro et al. (2011). We supplement this baseline dataset with grades from high school transcripts, risky behaviors at young ages, and later-life outcomes that were not previously available, such as physical health at age 40. Table A1 provides an overview of how our base sample is constructed, and how many observations are lost at each point.

Table A1: NLSY79 Data Set Construction and Effect of Deletions

Observations	Details
3,002	Core representative male NLSY population
2,975	Require schooling defined (GED or HS) for 12 years completed
2,905	Not employed by military
2,763	Not enrolled in education at 30 years old
2,242	Require no missing education, covariates, ASVAB, Rosenberg, and, instruments (Heckman et al. (2006) sample)

A.3 The Relationship Between Our Continuation Values and Those Estimated in the Dynamic Discrete Choice Literature

Dynamic discrete choice models greatly facilitate the interpretation of intertemporal choices and their consequences. Consider a basic dynamic human capital model analyzed by Keane and Wolpin (1997). Assume risk-neutral agents who have a finite choice set with N alternatives over a finite decision horizon $(\underline{a}, \overline{a})$. Let $B_n(a) = 1$ if alternative n is chosen at age a and zero otherwise. Let $R_n(a)$ be the current flow reward at age a from alternative n. The current reward per period at any age a is

$$R(a) = \sum_{n=1}^{N} \underbrace{R_n(a)}_{\substack{\text{per}\\\text{period}\\\text{reward}\\\text{from}\\\text{choice } n}} \underbrace{B_n(a)}_{\substack{\text{choice}\\\text{indicator}}} .$$

Denote an individual's state at age a by $\boldsymbol{H}(a)$. Assume a discount factor δ. The value function is

$$V\left(\boldsymbol{H}(a), a\right) = \max_{B_n(\tau) \in \mathcal{B}(a)} E\left[\sum_{\tau=a}^{\overline{a}} \delta^{\tau-a} \sum_{n=1}^{N} R_n(\tau) B_n(\tau) \mid \boldsymbol{H}(a)\right],$$

where $\mathcal{B}(a)$ is the set of feasible current and future choices at age a. The alternative-specific functions, $V_n\left(\boldsymbol{H}(a), a\right)$, can be written as

$$\begin{aligned} V_n\left(\boldsymbol{H}(a), a\right) &= R_n\left(\boldsymbol{H}(a), a\right) \\ &+ \delta E \underbrace{\left[V\left(\boldsymbol{H}(a+1), a+1\right) \mid \boldsymbol{H}(a), B_n(a) = 1\right]}_{\text{Continuation Value}} \end{aligned}$$

for $a < \overline{a}$, where $V_n\left(\boldsymbol{H}(\overline{a}), \overline{a}\right) = R_n\left(\boldsymbol{H}(\overline{a}), \overline{a}\right)$, the reward in state n at age a for a person with history $H(\overline{a})$. The decision rule is $B_n(a) = 1$ if $n = \underset{j \in \{1,\ldots,N\}}{\operatorname{argmax}} \{V_j(\boldsymbol{H}(a), a)\}$; $B_n(a) = 0$ otherwise.

Fully specified dynamic discrete choice models postulate agent preferences, constraints, and information sets, and can recover continuation values associated with each choice as well as option values that arise from moving up decision trees like those of Figure 1 in the text.[14] These benefits come at a price and many empirical economists reject the strong assumptions invoked in this and related literatures using dynamic economic models.

This paper takes a more agnostic and data-sensitive approach. We estimate a dynamic treatment effect model that captures some essential features of dynamic discrete choice models, but does not impose specific functional forms and decision models and assumptions about distributions of unobservables. We estimate *ex-post* approximations to the continuation values of a dynamic discrete choice model.

A.4 Precise Parameterization of the Model and Our Likelihood

This section presents more details on how the model is parameterized and estimated.

A.4.1 Parameterization of the Model

Following a well-established tradition in the literature,[15] in this paper we approximate I_j using a linear-in-the-parameters model:

$$I_j = \boldsymbol{Z}_j'\boldsymbol{\beta}_j + \boldsymbol{\theta}'\boldsymbol{\gamma}_j - \nu_j,\ j \in \{0, \dots, \overline{s} - 1\}, \tag{A.1}$$

where $\boldsymbol{Z}_j$ 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, and $\boldsymbol{\theta}$ 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 measurements. $\boldsymbol{\theta}$ plays an important role

[14] See, e.g., Eisenhauer et al. (2015) for estimates of a structural model of schooling with option values and continuation values.

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

in our model. Along with the observed variables, it generates dependence among schooling choices and outcomes. ν_j represents an idiosyncratic error term assumed to be independent across agents and states: $\nu_j \perp\!\!\!\perp (\boldsymbol{Z}_j, \boldsymbol{\theta})$, where "$\perp\!\!\!\perp$" denotes statistical independence.

Outcomes are also approximated by a linear-in-the-parameters model.

$$\tilde{Y}_s^k = (\boldsymbol{X}_s^k)'\boldsymbol{\beta}_s^k + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k + \omega_s^k, \tag{A.2}$$

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

A.4.2 Measurement System for Unobserved Endowments $\boldsymbol{\theta}$

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 outcomes.[16] 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 non-cognitive endowments 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 identify selection-bias-free estimates of causal effects and model parameters. We do not directly measure $\boldsymbol{\theta}$, and instead, we proxy it and correct for the effects of measurement error on the proxy. 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 under the conditions stated in Heckman and Navarro (2007).

Let θ^C and θ^{SE} denote the levels of cognitive and socio-emotional endowments and suppose

[16] See Heckman (2008).

$\boldsymbol{\theta} = (\theta^C, \theta^{SE})$. We allow θ^C and θ^{SE} to be correlated. Let $t^C_{m,s}$ be the m^{th} cognitive test score and $t^{C,SE}_{m,s}$ the m^{th} measure influenced by both cognitive and socio-emotional 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:

$$t^C_{m,s} = (\boldsymbol{X}^C_{m,s})'\boldsymbol{\beta}^C_{m,s} + \theta^C \alpha^C_{m,s} + e^C_{m,s}, \tag{A.3}$$

$$t^{C,SE}_{m,s} = (\boldsymbol{X}^{C,SE}_{m,s})'\boldsymbol{\beta}^{C,SE}_{m,s} + \theta^C \tilde{\alpha}^C_{m,s} + \theta^{SE} \tilde{\alpha}^{SE}_{m,s} + e^{C,SE}_{m,s}. \tag{A.4}$$

The structure assumed in Equations (A.3) and (A.4) is identified even when allowing for correlated factors, if we have one measure that is a determinant of cognitive endowments ($t^C_{m,s}$) and at least four measures that load on both cognitive ability and socio-emotional ability, and conventional normalizations are assumed.[17] In the main text we report results from models that use measurements to proxy $\boldsymbol{\theta}$. Let $H^m_{i,s}$ be an indicator for if an individual i took test t at schooling level s.

A.4.2.1 Specification of the Measurement System When estimating the factor model, we must make normalizations and exclusion restrictions. There is no precise method for determining these restrictions. As laid out below, we use a collection of empirical evidence and theory for determining our measurement system.

Factors have no natural scale. To address this, we normalize one loading for each factor to unity. This normalization does not affect the relative loadings of the two factors, but rather determines the units in which the factors are measured. We normalize the measure that has the largest correlation with the other measures. In the case of our paper, we normalize the cognitive loading to one for the arithmetic reasoning ASVAB measure, and we normalize the socio-emotional loading to one for the language arts grade measure. Switching the normalization to the loadings on other measures has no substantive effect on the results.

[17] See, e.g., the discussion in Williams (2011) and Anderson and Rubin (1956). 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 et al. (2010).

Following Heckman et al. (2006), the model imposes that the ASVAB measures do not load on socio-emotional factors. If any particular ASVAB score is excluded, it does not substantively change the analysis. Course grades are assumed to load on both the cognitive and socio-emotional factors. As discussed in the main paper, this assumption is also made by Duckworth and Seligman (2005) and Borghans et al. (2011), who both find that grades are largely determined by endowments other than cognitive ability.

As discussed above, the identification strategy used in the paper requires one measure that loads exclusively on cognitive ability. We assume ASVAB tests only measure cognition. Subject-specific 9th grade GPA, educational choices, and early risky behavior are assumed to depend on both factors.

We include violent behavior, smoking regularly by age 15, drinking regularly by age 15, ever smoking marijuana by age 15, and sexual intercourse by age 15 as early "outcomes" in our model. These do not inform the cognitive or socio-emotional factor, but provide a robustness check of our interpretation of our factors and aid in interpretation.

A.4.3 Likelihood

We estimate our model in two stages using maximum likelihood. The measurement system, 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 et al. (2004), and correct estimated factor distributions for the causal effect of choices on measurements by jointly estimating the choice and measurement equations in the first stage. The distribution of the latent factors is estimated only using data on educational choices and measurements. This allows us to interpret the factors as cognitive and socio-emotional endowments. It links our estimates to an emerging literature on the economics of personality and psychological traits, but the link is not strictly required if we only seek to control for selection in schooling choices and do not seek to identify the system of measurement equations presented in the text. We do not use the final outcome system to estimate the distribution of factors, thus avoiding tautologically strong predictions of outcomes from the system of estimated factors.

For convenience, we repeat the definitions from Section 2. Let $\mathcal{J}$ denote the set of possible terminal states. Let $D_j \in \mathcal{D}$ be the set of possible transition decisions that can be taken by the individual over the decision horizon. Let $\mathcal{S}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$. Define $\overline{s}$ as the highest attainable element in $\mathcal{S}$. $Q_j = 1$ indicates that an agent *gets to* decision node j. $Q_j = 0$ if the person never gets there. The history of nodes visited by an agent can be described by the collection of the Q_j such that $Q_j = 1$. To ensure consistent notation, we define $Q_0 := 1$. $\boldsymbol{Y}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{M}_i$ are vectors of individual i's outcomes, educational decisions, and measurements of endowments, respectively. $\boldsymbol{Z}$ is a vector of observed determinants of decisions, $\boldsymbol{X}$ is a vector of observed determinants of outcomes, and $\boldsymbol{\theta}$ is the vector of unobserved endowments. The $\boldsymbol{Z}$ can include all variables in $\boldsymbol{X}$. When instrumental variable methods are used to identify components of the model, it is assumed that there are some variables in $\boldsymbol{Z}$ not in $\boldsymbol{X}$.

Assuming independence across individuals (denoted by i), the likelihood is:

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

where $f(\cdot)$ denotes a probability density function. The last step is justified from the assumptions (A-1a)–(A-1g).

For the first stage, the sample likelihood is

$$
\begin{aligned}
\mathcal{L}^1 &= \prod_i \int_{\overline{\theta}\in\Theta} f(\boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) f_\theta(\overline{\boldsymbol{\theta}})\, d\overline{\boldsymbol{\theta}} \\
&= \prod_i \int_{\overline{\theta}\in\Theta} \left[\prod_{j\in\mathcal{S}\setminus\{\overline{s}\}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \\
&\qquad \times \left[\prod_{m=1}^{N_M} \prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{m,s})^{H^m_{i,s}} \right] f_\theta(\overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_\theta)\, d\overline{\boldsymbol{\theta}},
\end{aligned}
$$

where we integrate over the distributions of the latent factors. H^m_s is an indicator for the level of the choice variable at the time the measurement m is taken, and is equal to one if the individual had attained s at the time of the measurement, and zero otherwise. Let $\mathcal{S}^M$ denote the set of possible states at the time of the measurement. The goal of the first stage is to secure estimates of $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$, and $\boldsymbol{\gamma}_\theta$, where $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$, and $\boldsymbol{\gamma}_\theta$ are the parameters for the educational decision models, the measurement models, and the factor distribution, respectively. We assume that the idiosyncratic shocks are mean zero normal variates.

We approximate the factor distribution using a mixture of normals.[18] We define the index, ℓ, for each mixture, where $f_\theta(\boldsymbol{\theta}; \boldsymbol{\gamma}_\theta) = \sum_\ell \rho_\ell f^\ell_\theta(\boldsymbol{\theta}; \boldsymbol{\gamma}^\ell_\theta)$. The weights for each mixture are ρ_ℓ and they must satisfy $\sum_\ell \rho_\ell = 1$. $f^\ell_\theta(\boldsymbol{\theta}; \boldsymbol{\gamma}^\ell_\theta)$ is the PDF for mixture ℓ. Since the mean of the overall factor distribution is not identified, we also require that $E[\boldsymbol{\theta}] = \mathbf{0}$ which places

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

constraints on the mixture parameters $\boldsymbol{\gamma}_\theta^\ell$. The log-likelihood can be rewritten as

$$\begin{aligned}\log \mathcal{L}^1 = \sum_i \log \int_{\overline{\theta}\in\Theta} & \left[\prod_{j\in\mathcal{S}\setminus\overline{s}} f(\boldsymbol{D}_{i,j}|\boldsymbol{Z}_{i,j},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_j)^{Q_{i,j}}\right] \\ & \times \left[\prod_{m=1}^{N_M}\prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s}|\boldsymbol{X}_{i,m,s},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_{m,s})^{H_{i,s}^m}\right] \times \left[\sum_\ell \rho_\ell f_\theta^\ell(\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_\theta^\ell)\right] d\overline{\boldsymbol{\theta}} \\ = \sum_i \log & \left\{\sum_\ell \rho_\ell \int_{\overline{\theta}\in\Theta}\left[\prod_{j\in\mathcal{S}\setminus\overline{s}} f(\boldsymbol{D}_{i,j}|\boldsymbol{Z}_{i,j},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_j)^{Q_{i,j}}\right]\right. \\ & \left.\left[\prod_{m=1}^{N_M}\prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s}|\boldsymbol{X}_{i,m,s},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_{m,s})^{H_{i,s}^m}\right] f_\theta^\ell(\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_\theta^\ell)d\overline{\boldsymbol{\theta}}\right\}.\end{aligned}$$

We use Gauss-Hermite quadrature to numerically evaluate the integral. Although there are a number of ways to numerically evaluate an integral, one advantage of Gaussian quadrature is that it gives analytical expressions for the integral. Analytical expressions for the gradient and Hessian can then be calculated, which allows for the use of efficient second-order optimization routines. Since the models are very smooth, a second-order optimization strategy leads to faster convergence. Given that we are using a mixture of normals, $f_\theta^\ell(\boldsymbol{\theta};\boldsymbol{\gamma}_\theta^\ell) = \phi(\boldsymbol{\theta};\boldsymbol{\mu}_\theta^\ell,\boldsymbol{\sigma}_\theta^\ell)$ is a multivariate normal, where we assume for now that the components are independent. This assumption can easily be relaxed, but keeping it simplifies notation. The Gauss-Hermite quadrature rule is $\int f(v)e^{-v^2}dv = \sum_n \lambda_n f(v_n)$, where the weights, λ_n, and nodes, v_n, are defined by the quadrature rule depending on the number of points used (Judd, 1998).[19] Applying the Gauss-Hermite rule and making a change of variables ($\overline{\boldsymbol{\theta}} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell$),[20] we can rewrite the likelihood as

[19] We use 16 quadrature points. Using 32 points did not substantively change any of our results.
[20] $\circ$ is the Hadamard or entrywise product.

$$\log \mathcal{L}^1 = \sum_i \log \left\{ \sum_\ell \rho_\ell \sum_{n_1} \lambda_{n_1} \sum_{n_2} \lambda_{n_2} \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \right.$$
$$\left. \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell; \boldsymbol{\gamma}_{m,s})^{H_{i,s}^m} \right] \right\},$$

where $\boldsymbol{v_n} = (v_{n_1}, v_{n_2})$ represents the vector of nodes. Multivariate normal variables with correlated components can be rewritten as the sum of independent standard normal variables, and then one can use the same procedure.

The goal of the first stage is then to maximize $\log \mathcal{L}^1$ and obtain estimates $\hat{\boldsymbol{\gamma}}_j$, $\hat{\boldsymbol{\gamma}}_{m,s}$, $\hat{\boldsymbol{\sigma}}_\theta^\ell$, $\hat{\boldsymbol{\mu}}_\theta^\ell$, and $\hat{\rho}_\ell$ for $j \in J^{MS}$. If a density $f(\cdot)$ cannot be calculated, either because of missing data or because that model does not apply to individual i,[21] then $f(\cdot) = 1$.

One can think of the inner brackets as the PDF of $\boldsymbol{\theta}$ for each individual i. This is useful in two respects. First, we can now predict the factor scores ($\hat{\boldsymbol{\theta}}_i$) via maximum likelihood where the likelihood for each individual i is

$$\mathcal{L}_i^\theta = \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}} \right] \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_{m,s})^{H_{i,s}^m} \right].$$

Secondly, we can correct for measurement error in the outcome equations by integrating over the PDF of the latent factor. The likelihood for the outcome equations is

$$\log \mathcal{L}_k^2 = \sum_i \log \left\{ \sum_\ell \rho_\ell \sum_{n_1} \lambda_{n_1} \sum_{n_2} \lambda_{n_2} \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}} \right] \right.$$
$$\times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \hat{\boldsymbol{\gamma}}_{m,s})^{H_{i,s}^m} \right]$$
$$\left. \times \left[\prod_{s \in \mathcal{S}} f(\boldsymbol{Y}_{i,s}^k | \boldsymbol{X}_{i,k,s}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \boldsymbol{\gamma}_{s,k})^{H_{i,s}} \right] \right\},$$

[21] For example, the individual i is a high school dropout and the model corresponds to the graduate college decision.

where $H_{i,s}$ is an indicator for the highest level of schooling attained by individual i. The goal of the second stage is to maximize $\log \mathcal{L}_k^2$ and obtain estimates $\hat{\boldsymbol{\gamma}}_{s,k}$. Since outcomes ($\boldsymbol{Y}_s^k$) are independent from the first stage outcomes conditional on $\boldsymbol{X}, \boldsymbol{\theta}$ and we impose no cross-equation restrictions, we obtain consistent estimates of the parameters for the adult outcomes. Standard errors and confidence intervals are calculated by estimating two hundred bootstrap samples for the combined stages.

A.5 Goodness of Fit

This section tests the goodness of fit of our various measurement, education, and outcome equations. For continuous models we compare means and standard deviations, while for discrete outcomes we compare proportions. Table A7 jointly tests for equality of means in the outcomes for 16 unique sub-populations.

Table A2: Goodness of Fit - Schooling Choice

Schooling Level	Data	Model	p-value
High School Dropout	0.131	0.122	0.980
High School Graduate	0.370	0.377	0.989
Some College	0.168	0.176	0.982
College Graduate	0.230	0.222	0.986

Notes: The simulated data (Model) contains one million observations generated from the model estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is *Model=Data*.

Table A3: Goodness of Fit - Early Risky and Reckless Behavior

Outcome	Actual	Model	p-value[a]
Early Marijuana[c]	0.338	0.338	0.999
Early Daily Smoking[c]	0.187	0.186	0.999
Early Drinking[c]	0.188	0.188	0.999
Early Intercourse[c]	0.163	0.161	0.994
Early Reckless (9th–11th)[b]	0.607	0.599	0.987
Early Reckless (12th)[b]	0.533	0.541	0.988

Notes: The simulated data (Model) contains one million observations generated from the model estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is that the model fits the data. (b) The reckless and violent variables are taken from the NSLY 1980 Illegal Activities Supplement. (c) Early is defined as engaging in risky behavior before 15 years old.

Table A4: Goodness of Fit - ASVAB and Grade Models

	Mean		Std Dev		
	Data	Model	Data	Model	p-value
ASVAB Tests					
Arithmetic Reasoning (< 12)	-0.291	-0.354	0.932	0.898	0.035
Word Knowledge (< 12)	-0.448	-0.530	1.082	1.057	0.017
Paragraph Comprehension (< 12)	-0.513	-0.588	1.180	1.166	0.049
Numerical Operations (< 12)	-0.519	-0.574	0.963	0.927	0.074
Math Knowledge (< 12)	-0.320	-0.389	0.887	0.835	0.015
Coding Speed (< 12)	-0.599	-0.643	0.782	0.767	0.075
Arithmetic Reasoning ($= 12$)	0.196	0.186	0.862	0.823	0.720
Word Knowledge ($= 12$)	0.132	0.126	0.778	0.735	0.837
Paragraph Comprehension ($= 12$)	0.039	0.029	0.796	0.751	0.699
Numerical Operations ($= 12$)	-0.012	-0.020	0.890	0.848	0.787
Math Knowledge ($= 12$)	0.001	-0.022	0.812	0.745	0.417
Coding Speed ($= 12$)	-0.163	-0.167	0.773	0.749	0.866
Arithmetic Reasoning (> 12)	0.942	0.905	0.665	0.643	0.258
Word Knowledge (> 12)	0.744	0.754	0.328	0.307	0.552
Paragraph Comprehension (> 12)	0.636	0.622	0.324	0.314	0.377
Numerical Operations (> 12)	0.580	0.560	0.589	0.571	0.475
Math Knowledge (> 12)	0.975	0.947	0.736	0.720	0.438
Coding Speed (> 12)	0.474	0.460	0.654	0.639	0.665
9th Grade GPA					
GPA Language	-0.117	-0.175	0.969	0.973	0.012
GPA Social Sciences	-0.012	-0.074	0.985	0.993	0.018
GPA Science	0.026	-0.017	0.955	0.939	0.085
GPA Math	-0.011	-0.050	0.977	0.975	0.083

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.The numbers inside the parentheses describe the years of schooling at the time of the test. The ASVAB 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. (a) The p-values reported are from a T-test for the equivalence of the means where the null hypothesis is that $Actual = Model$.

Table A5: Goodness of Fit - Discrete Outcomes

Outcome	Actual	Model	p-value[a]
Smoking Age 30	0.385	0.387	0.997
High school dropouts	0.674	0.650	0.959
High school graduates	0.390	0.383	0.989
Some college	0.337	0.339	0.995
Four-year college graduate	0.146	0.166	0.955
Health Limits Work	0.227	0.226	0.997
High school dropouts	0.392	0.412	0.968
High school graduates	0.232	0.229	0.994
Some college	0.184	0.179	0.992
Four-year college graduate	0.091	0.099	0.980

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is that the model predictions fits the data.

Table A6: Goodness of Fit - Continuous Outcomes

	Mean		Std Dev		
	Actual	Model	Actual	Model	p-value
Log Wages (30)	2.612	2.604	0.229	0.223	0.132
High school dropouts	2.291	2.247	0.135	0.130	0.000
High school graduates	2.531	2.528	0.184	0.182	0.637
Some college	2.665	2.677	0.207	0.200	0.283
Four-year college graduate	2.932	2.949	0.188	0.186	0.039
PV Log Wages (30)	12.315	12.317	0.397	0.395	0.876
High school dropouts	11.787	11.681	0.366	0.391	0.000
High school graduates	12.275	12.275	0.273	0.262	0.983
Some college	12.422	12.432	0.257	0.255	0.499
Four-year college graduate	12.764	12.817	0.266	0.272	0.000

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) The p-values reported are from a T-test for the equivalence of the means where the null hypothesis is that the model predictions fits the data.

Table A7: χ^2 Test for Equality of the Means Across Sub-Populations

	p-value
Log Wages (30)	0.44
Log PV Wage Income (30)	0.07
Health Limits Work	0.40
Smoking (30)	0.06

Notes: This table jointly tests if the observed and simulated outcome means are equal for 16 unique sub-populations. Sub-populations are the unique groups defined by the binary variables: white, Southern residence at age 14, family income greater than $20,500 in 1979, and mother's highest grade completed is less than 12. The reported p-value is for the χ^2-test against the null hypothesis that the means are equal for the observed and simulated data.

A.6 Estimated Model Parameters

This section documents the estimated model parameters from our maximum likelihood procedure. The first table shows the educational choice models while the remaining tables show the various outcome models conditional on final educational attainment.

Table A8: Estimates for Schooling Choice Model

Variable	D_1: Graduate HS vs. Drop out of HS		$D_{5'}$: **GED** vs. **HS Dropout**		D_2: Enroll College vs. **HS Graduate**		D_3: **Four-Year College Degree** vs. **Some College**	
	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.
Black	0.081	0.129	-0.120	0.178	0.144	0.140	0.014	0.196
Hispanic	0.648	0.179	-0.083	0.252	0.631	0.175	0.421	0.255
Broken Home	-0.484	0.101	-0.241	0.140	-0.049	0.103	-0.283	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	0.095	0.021	0.100	0.027
Father's Education	0.065	0.016	0.038	0.026	0.126	0.015	0.103	0.019
Age	-0.161	0.489	0.027	0.771	0.158	0.469	-0.626	0.485
Age^2	0.005	0.013	-0.002	0.020	-0.003	0.012	0.016	0.012
Family Income	0.020	0.005	0.018	0.008	0.012	0.004	0.013	0.005
Intercept	0.728	4.754	-1.612	7.501	-4.594	4.549	3.268	4.657
Urban	-0.180	0.097	0.511	0.161	0.089	0.095		
South	-0.414	0.121	0.335	0.195	0.159	0.118		
West	-0.403	0.127	0.213	0.214	-0.178	0.128		
Northeast	0.207	0.126	0.141	0.225	0.384	0.111		
Local Unemployment	2.454	2.388	5.721	3.753	4.526	2.330		
Local Average Unemployment$^{(a)}$	-10.464	4.805	-2.618	7.326	-6.061	4.546	-5.386	4.925
GED Passrate			0.000	0.068				
Tuition at Local Four-Year College					-0.256	0.062	-0.016	0.086
Local College Present					0.151	0.080	0.059	0.106
Local Unemployment							-0.044	1.909
Urban							0.048	0.136
South							-0.088	0.154
West							-0.442	0.182
Northeast							0.068	0.159
Cognitive	0.815	0.100	1.010	0.138	0.847	0.097	0.834	0.117
Socio-Emotional	0.984	0.090	0.171	0.145	0.496	0.082	0.596	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 average 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. Presence of a local college affects the decision to enroll and graduate from College.

Table A9: Estimates for Log Wages at Age 30

Variables	Full		HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.212	0.032	-0.200	0.068	-0.279	0.089	-0.229	0.052	-0.247	0.081	-0.066	0.079
Hispanic	-0.045	0.041	-0.178	0.077	-0.049	0.143	-0.004	0.065	-0.136	0.103	-0.009	0.109
Broken Home	-0.008	0.023	-0.022	0.047	0.139	0.073	0.064	0.038	-0.082	0.059	-0.063	0.056
Number of Siblings	-0.007	0.004	-0.006	0.009	0.009	0.014	-0.007	0.007	0.008	0.011	0.006	0.010
Mother's Education	0.013	0.005	0.004	0.011	0.007	0.017	-0.001	0.008	0.014	0.012	0.013	0.009
Father's Education	0.012	0.003	0.008	0.009	0.019	0.012	0.012	0.006	-0.006	0.009	0.006	0.007
Age	0.025	0.086	0.151	0.213	0.331	0.329	-0.136	0.133	0.190	0.211	0.097	0.175
Age^2	-0.000	0.002	-0.004	0.006	-0.007	0.009	0.004	0.003	-0.005	0.005	-0.002	0.005
Family Income	0.007	0.001	0.008	0.003	0.008	0.004	0.007	0.002	0.008	0.002	0.005	0.001
Intercept	1.709	0.810	0.586	2.015	-1.692	3.078	3.251	1.255	0.451	1.982	1.363	1.650
Local Unemployment	-0.367	0.482	-1.023	1.192	0.909	1.605	0.511	0.759	-1.520	1.199	-0.986	1.050
Northeast	0.141	0.027	0.291	0.073	0.125	0.115	0.069	0.041	0.129	0.070	0.183	0.050
South	-0.010	0.024	0.078	0.058	0.024	0.089	-0.046	0.038	0.016	0.063	-0.025	0.049
West	0.030	0.028	0.059	0.076	0.005	0.107	0.046	0.045	0.062	0.066	-0.001	0.056
Urban	0.119	0.023	0.055	0.054	0.108	0.084	0.098	0.033	0.117	0.060	0.139	0.056
Cognitive	0.199	0.015	0.095	0.052	0.147	0.059	0.157	0.024	0.040	0.042	0.233	0.042
Socio-emotional	0.054	0.018	-0.056	0.054	0.078	0.072	-0.049	0.029	0.033	0.051	0.057	0.045
Std. Error	0.399	0.006	0.324	0.015	0.424	0.022	0.385	0.010	0.411	0.016	0.387	0.013
N	1991		235		183		757		340		476	

Notes: Table shows estimated coefficients for the full population and conditional on each final schooling level.

Table A10: Estimates for Log PV Wages

Variables	Full		HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.363	0.038	-0.745	0.089	-0.351	0.114	-0.345	0.060	-0.318	0.091	-0.104	0.101
Hispanic	0.039	0.051	-0.152	0.113	-0.076	0.192	0.200	0.080	-0.315	0.109	0.315	0.137
Broken Home	-0.117	0.028	-0.079	0.065	-0.044	0.095	-0.042	0.044	-0.122	0.066	-0.120	0.070
Number of Siblings	-0.014	0.005	-0.037	0.012	0.031	0.018	-0.005	0.008	-0.003	0.014	-0.003	0.013
Mother's Education	0.030	0.006	0.026	0.015	0.032	0.022	0.018	0.010	0.015	0.014	0.036	0.012
Father's Education	0.016	0.004	0.016	0.012	0.021	0.016	0.019	0.007	-0.010	0.010	0.003	0.009
Age	-0.135	0.098	-0.269	0.268	-0.299	0.379	0.012	0.143	0.025	0.221	-0.229	0.195
Age^2	0.004	0.003	0.007	0.007	0.008	0.010	-0.000	0.004	-0.001	0.006	0.006	0.005
Family Income	0.011	0.001	0.018	0.004	0.017	0.005	0.012	0.002	0.008	0.002	0.008	0.002
Intercept	12.817	0.941	14.015	2.571	13.709	3.621	11.453	1.382	11.967	2.111	14.095	1.879
Urban	0.084	0.027	0.216	0.070	0.038	0.111	0.045	0.039	0.175	0.062	0.066	0.060
South	0.039	0.029	0.065	0.076	0.158	0.115	0.024	0.045	0.029	0.069	0.031	0.058
West	-0.038	0.036	-0.144	0.101	-0.157	0.137	-0.027	0.052	0.116	0.076	-0.143	0.078
Northeast	0.107	0.032	0.161	0.097	-0.001	0.147	0.035	0.048	0.101	0.076	0.158	0.058
Cognitive	0.276	0.018	0.428	0.072	0.408	0.075	0.180	0.029	0.079	0.046	0.231	0.053
Socio-emotional	0.086	0.022	0.018	0.077	0.018	0.093	-0.089	0.035	-0.010	0.056	0.137	0.057
Std. Error	0.495	0.008	0.452	0.021	0.581	0.029	0.448	0.012	0.456	0.018	0.464	0.016
N	1986		254		204		744		333		451	

Notes: Table shows estimated coefficients for the full population and conditional on each final schooling level.

Table A11: Estimates for Smoking at Age 30

Variables	Full		HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.016	0.106	-0.199	0.277	-0.106	0.271	0.216	0.180	-0.349	0.288	-0.082	0.359
Hispanic	-0.597	0.146	-0.941	0.352	-1.298	0.467	-0.296	0.245	-0.413	0.342	0.519	0.431
Broken Home	0.254	0.078	0.316	0.209	-0.225	0.222	0.037	0.129	0.623	0.199	-0.008	0.253
Number of Siblings	0.029	0.015	0.103	0.043	-0.020	0.045	-0.003	0.025	0.053	0.038	-0.019	0.043
Mother's Education	-0.038	0.016	0.004	0.046	-0.072	0.054	0.018	0.028	0.031	0.042	-0.076	0.040
Father's Education	0.013	0.012	0.054	0.040	0.031	0.038	0.046	0.022	0.008	0.029	0.076	0.033
Age	0.130	0.269	0.237	0.821	0.393	0.910	-0.342	0.417	0.543	0.675	-0.297	0.685
Age^2	-0.003	0.007	-0.004	0.021	-0.008	0.024	0.009	0.011	-0.014	0.017	0.009	0.018
Family Income	-0.007	0.003	0.013	0.013	-0.006	0.012	-0.009	0.005	0.004	0.007	-0.003	0.005
Intercept	-1.526	2.593	-4.196	7.853	-3.489	8.632	2.421	4.019	-6.542	6.445	1.351	6.608
Northeast	-0.131	0.093	0.072	0.327	-0.323	0.361	0.132	0.139	-0.281	0.239	-0.199	0.215
South	0.017	0.080	0.023	0.246	-0.431	0.281	0.119	0.129	-0.120	0.208	0.010	0.198
West	-0.061	0.096	-0.291	0.303	-0.061	0.330	-0.212	0.155	0.018	0.217	-0.425	0.267
Urban	0.038	0.077	-0.103	0.225	0.147	0.257	0.043	0.113	0.112	0.196	0.421	0.286
Cognitive	-0.337	0.050	-0.340	0.229	-0.385	0.183	-0.023	0.083	-0.068	0.135	-0.233	0.181
Socio-emotional	-0.472	0.062	-0.441	0.233	-0.165	0.215	-0.132	0.103	-0.153	0.164	-0.361	0.195
N	1882		236		191		708		315		432	

Notes: Table shows estimated coefficients for the full population and conditional on each final schooling level.

Table A12: Estimates for Health Limits Work

Variables	Full		HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	0.162	0.098	0.361	0.228	0.033	0.242	0.102	0.174	0.173	0.269	0.142	0.321
Hispanic	-0.119	0.132	-0.178	0.292	-0.352	0.391	0.129	0.214	0.204	0.329	-0.783	0.612
Broken Home	0.129	0.075	0.080	0.173	-0.061	0.199	-0.051	0.132	0.303	0.195	0.250	0.238
Number of Siblings	0.012	0.014	0.002	0.033	-0.032	0.039	-0.001	0.025	0.038	0.038	0.019	0.044
Mother's Education	-0.038	0.016	-0.009	0.040	-0.066	0.048	-0.016	0.027	-0.021	0.042	-0.015	0.043
Father's Education	-0.024	0.012	0.011	0.033	-0.033	0.034	-0.051	0.021	0.023	0.030	-0.003	0.031
Age	0.325	0.268	0.032	0.722	-0.389	0.803	0.869	0.436	0.205	0.675	-0.349	0.725
Age^2	-0.007	0.007	0.002	0.019	0.012	0.021	-0.021	0.011	-0.006	0.017	0.012	0.018
Family Income	-0.002	0.003	-0.012	0.011	-0.013	0.010	-0.000	0.006	0.003	0.007	0.004	0.005
Intercept	-3.811	2.593	-1.762	6.968	4.127	7.706	-8.930	4.214	-2.751	6.442	0.925	7.047
Urban	0.025	0.076	-0.165	0.191	-0.002	0.237	0.204	0.120	-0.280	0.187	-0.113	0.224
South	-0.032	0.080	0.137	0.205	-0.288	0.242	-0.053	0.136	-0.195	0.207	0.160	0.216
West	0.238	0.094	0.408	0.277	0.060	0.275	0.249	0.153	0.069	0.224	0.406	0.270
Northeast	-0.048	0.090	0.375	0.259	-0.218	0.305	0.034	0.140	-0.418	0.242	0.067	0.224
Cognitive	-0.340	0.049	-0.314	0.190	-0.269	0.158	-0.309	0.085	0.013	0.140	-0.347	0.175
Socio-emotional	-0.232	0.058	0.335	0.201	-0.168	0.195	-0.147	0.100	-0.146	0.168	-0.124	0.204
N	2242		293		229		829		376		515	

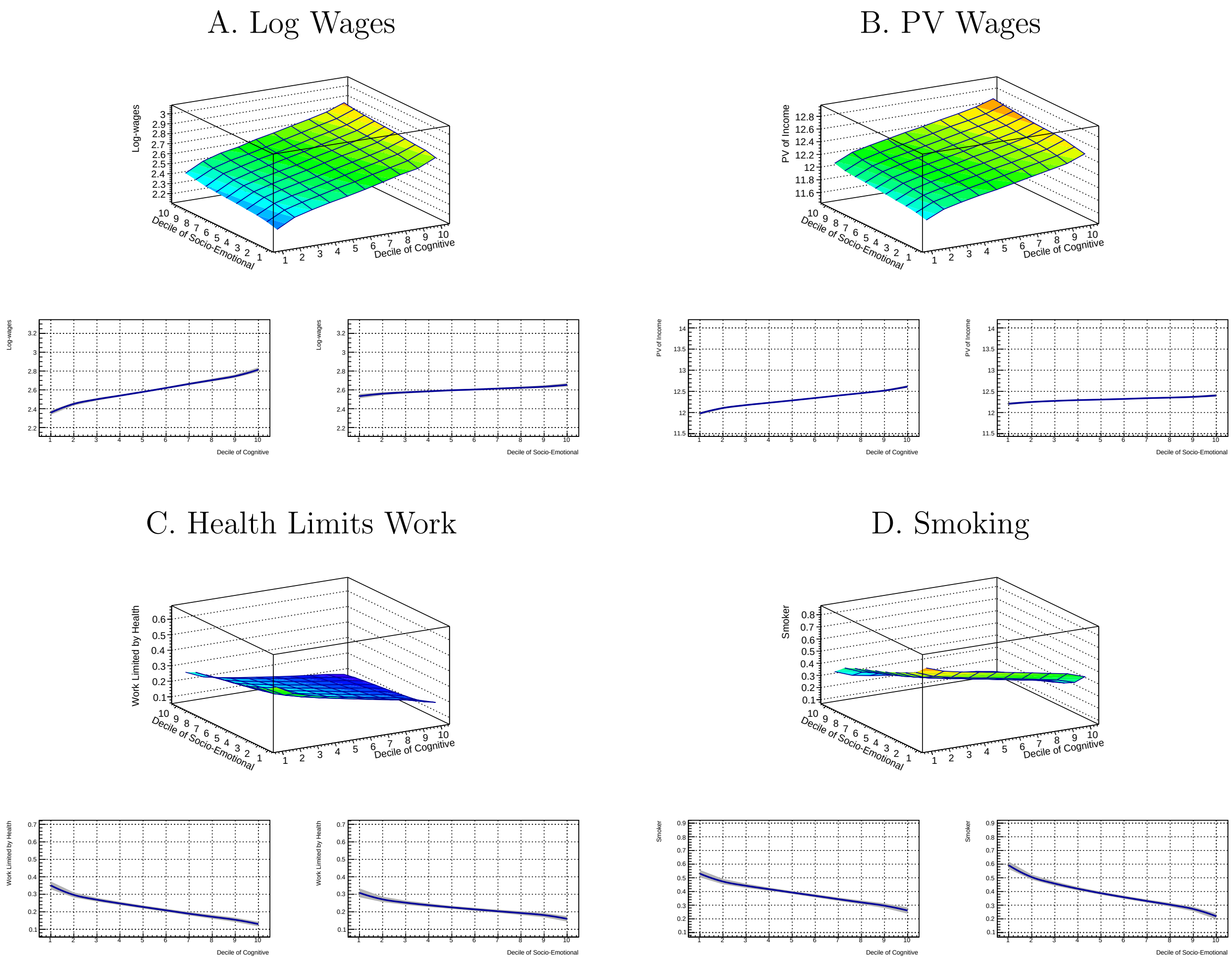
Notes: Table shows estimated coefficients for the full population and conditional on each final schooling level.

A.7 The Measurement of Endowments and Their Effects on Outcomes

A.7.1 The Role of Endowments on Later-Life Outcomes

The latent endowments have statistically significant effects on labor market and health outcomes. Figure A1 plots the effects of the latent endowments on log wages, log present value of log wage income, daily work limitations, and daily smoking. The cognitive endowment affects all four outcomes, while the effect of the socio-emotional 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 socio-emotional 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 deviation.

Figure A1: The Effect of Cognitive and Socio-Emotional Endowments

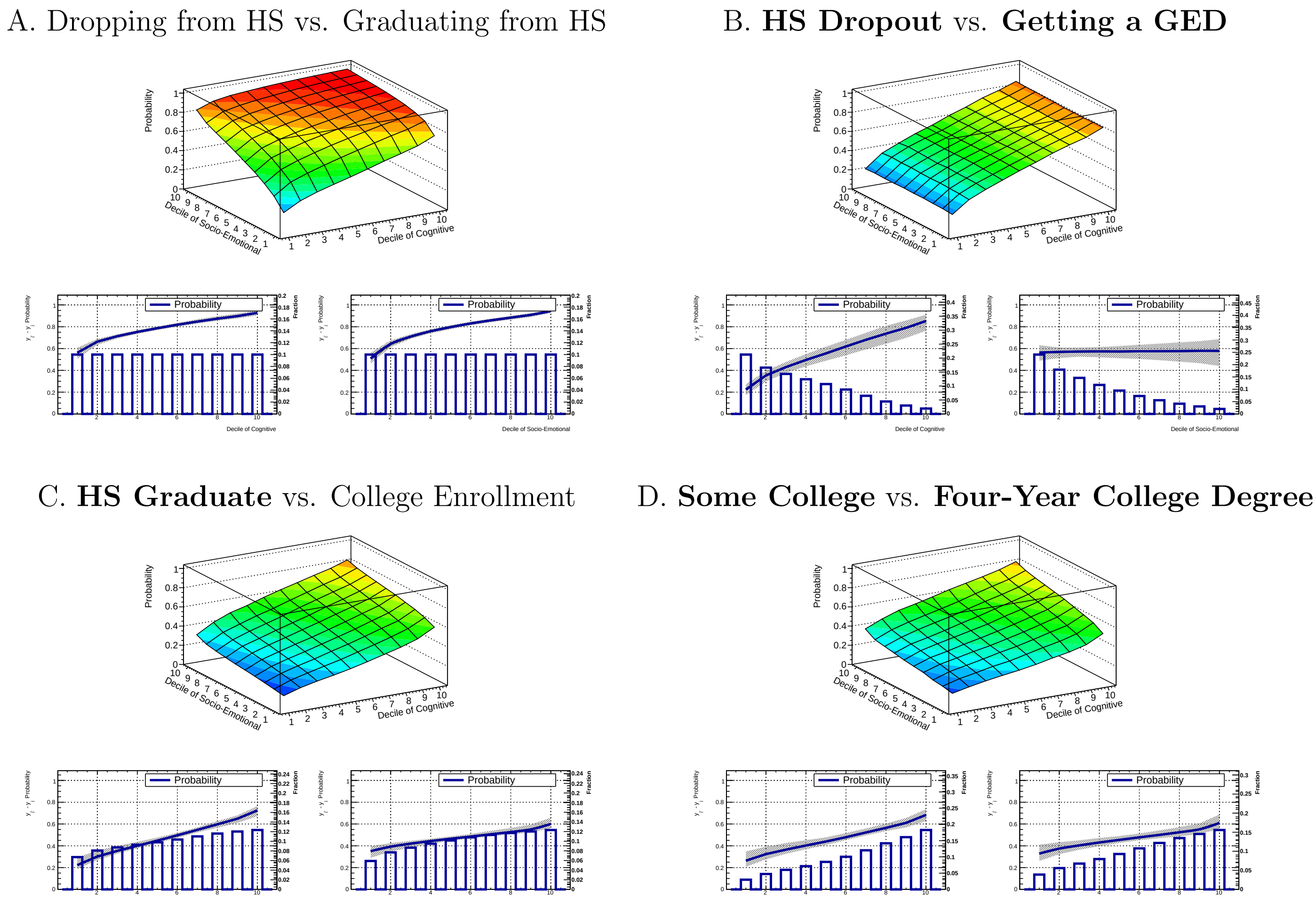


Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socio-emotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional 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 socio-emotional endowment.

A.7.2 Sorting on Unobserved Variables

Figure A2 presents the probabilities of making the indicated educational choice at various levels of agent latent endowments. Figure A4 shows the distribution of the factors by final schooling level. Individuals sort on both cognitive and socio-emotional endowments into increasing schooling levels. The only exception are the GEDs, who have cognitive ability distributions similar to terminal high school graduates but socio-emotional distributions similar to dropouts.

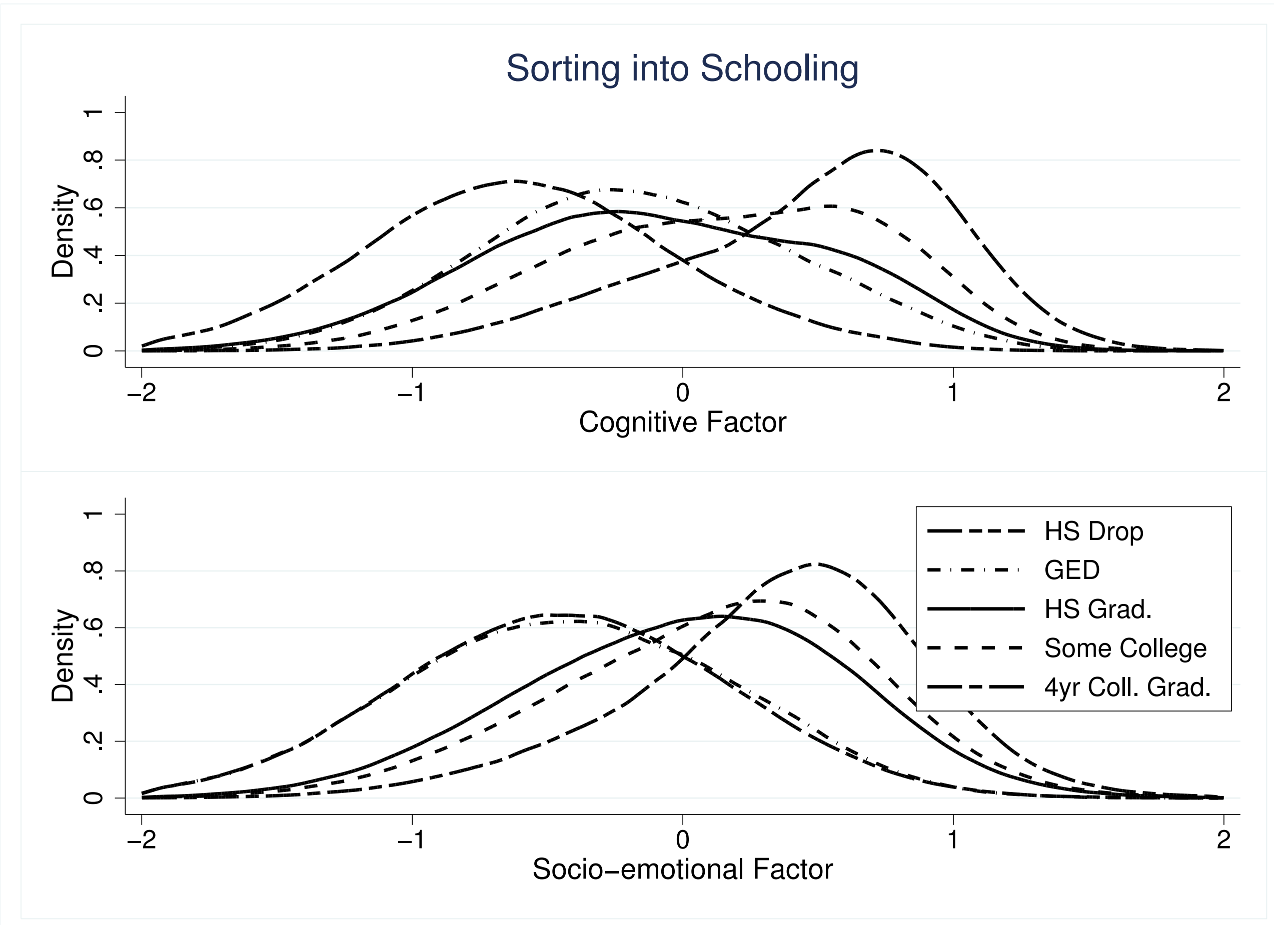
Figure A2: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters)



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 j and $j+1$, the first subfigure (top) presents $Prob(D_j = 0|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments. $Prob(D_j = 0|d^C, d^{SE})$ is computed for those who reach the decision node involving a decision between levels j and $j+1$. The bottom left subfigures present $Prob(D_j = 0|d^C)$ where the socio-emotional 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 j and $j+1$. The bottom right subfigures present $Prob(D_j = 0|d^{SE})$ for a given decile of socio-emotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$.

The estimates reveal clear evidence of sorting into education by both cognitive and socio-emotional endowments. At the same time, these endowments have significant impacts on adult outcomes. Together, these results imply strong selection biases in observed differences

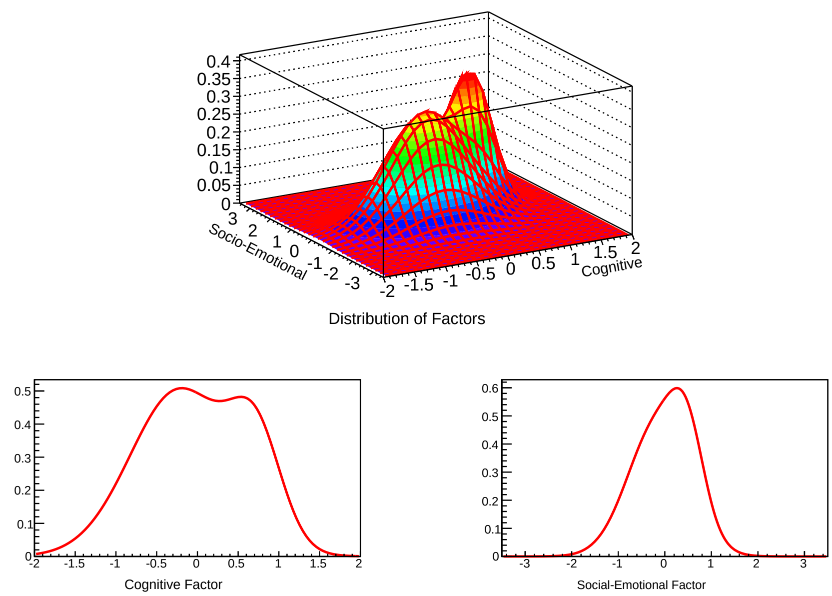
Figure A3: Distribution of Factors by Schooling Level



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

in outcomes by education level. This highlights the importance of accounting for observed and latent traits when estimating the causal impact of education.

Figure A4: Joint Distribution of Cognitive and Socio-Emotional Ability



$$\mathbf{\Sigma}_1 = \begin{pmatrix} 0.0971459 & 0 \\ 0 & 0.128441 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 0.367792 & 0 \\ 0 & 0.421223 \end{pmatrix},$$

Table A13: Means and Weights for Mixtures

	Mixture for Cognitive $\boldsymbol{\theta}$	Mixture for Non-Cognitive $\boldsymbol{\theta}$
μ_1	0.721206	-0.218251
μ_2	0.487487	-0.147523
Weight on First Component	0.232316	0.767684

A.7.3 Sorting on Observables

Table A14 shows the means and standard deviations of our controls by education level. There is sorting in nearly every background characteristic, except for age in 1980. The sorting is very strong in the cognitive factor, the socio-emotional factor, parental education, parental income, number of siblings, and growing up in a broken home.

Table A14: Educational Sorting on Observables

	Dropout	GED	High School	Some Coll.	Coll. Grad.
Cog	-0.651	-0.170	-0.119	0.162	0.522
	(0.535)	(0.651)	(0.667)	(0.641)	(0.533)
Socio-emotional	-0.677	-0.840	0.074	0.157	0.440
	(0.742)	(0.894)	(0.756)	(0.737)	(0.664)
Black	0.181	0.179	0.119	0.104	0.060
	(0.386)	(0.384)	(0.324)	(0.305)	(0.238)
Hispanic	0.113	0.074	0.064	0.080	0.031
	(0.317)	(0.263)	(0.245)	(0.271)	(0.174)
Broken Home	0.423	0.358	0.211	0.229	0.142
	(0.495)	(0.480)	(0.408)	(0.421)	(0.349)
Num. Siblings	4.181	3.777	3.379	2.923	2.538
	(2.640)	(2.585)	(2.177)	(2.122)	(1.807)
Mom's HGC	9.966	10.670	11.250	11.928	13.281
	(2.518)	(2.382)	(2.252)	(2.419)	(2.496)
Dad's HGC	9.768	10.577	11.115	12.312	14.290
	(3.099)	(2.997)	(2.872)	(3.251)	(3.397)
Fam. Inc. 1979	13.675	16.378	19.680	20.922	27.075
	(8.033)	(9.696)	(10.514)	(12.097)	(16.152)
Urban Age 14	0.771	0.790	0.715	0.755	0.804
	(0.421)	(0.408)	(0.452)	(0.430)	(0.397)
South Age 14	0.420	0.406	0.277	0.298	0.243
	(0.494)	(0.492)	(0.448)	(0.458)	(0.429)
Age in 1980	19.096	18.808	19.230	19.226	19.206
	(2.103)	(2.129)	(2.173)	(2.232)	(2.227)

Notes: Mean coefficients; SD in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There may be some concern that the socio-emotional factor is describing something like academic ability, since it is partly based on grades.[22] Table A15 provides additional

[22] "Risky and reckless behavior in 1979" is a binary measure of whether the person had ever done any

support for our interpretation of the socio-emotional factor. The table estimates the impact of observables and unobservables on early risky behavior, but is not used in estimating our factors. As we see, our non-cognitive factor plays an important role in each of the early risky outcomes. If the socio-emotional factor were measuring purely academic behavior, we would not expect it to be so predictive in explaining early risky behaviors.

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 include risky and reckless behavior in 1979 and find that the cognitive factor plays a small and statistically insignificant role, while the socio-emotional factor has a statistically significant impact.

Table A15: Early Outcomes: Estimates for "Early Risky Behaviors"

Variable	Tried Marijuana[a]		Daily Smoking[a]		Regular Drinking[a]		Intercourse[a]	
	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.
Black	-0.321	0.101	-0.341	0.112	-0.237	0.108	0.605	0.099
Hispanic	-0.160	0.125	-0.496	0.150	-0.010	0.130	-0.034	0.140
Broken Home	0.421	0.073	0.417	0.081	0.236	0.077	0.366	0.081
Number of Siblings	0.030	0.014	0.033	0.015	0.028	0.015	0.011	0.016
Mother's Education	0.011	0.015	-0.021	0.017	0.001	0.016	-0.022	0.017
Father's Education	-0.011	0.011	-0.036	0.013	-0.004	0.012	-0.027	0.013
Family Income	0.001	0.003	-0.003	0.003	-0.001	0.003	-0.003	0.003
Intercept	-0.165	2.473	-4.411	2.829	1.067	2.620	3.064	2.873
Age	0.022	0.257	0.384	0.293	-0.203	0.273	-0.406	0.298
Age^2	-0.003	0.007	-0.009	0.008	0.005	0.007	0.010	0.008
Urban	0.271	0.072	0.113	0.081	0.096	0.077	0.211	0.087
South	-0.110	0.067	-0.025	0.075	0.066	0.071	0.103	0.076
Cognitive	-0.102	0.048	-0.209	0.054	-0.137	0.052	-0.277	0.057
Socio-emotional	-0.616	0.060	-0.527	0.064	-0.288	0.061	-0.403	0.066
N	2239		2176		2231		2218	

Notes: 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.

A.8 Evidence on Equation (1) in the Text: Linearity of the Returns to Schooling

In this section, we use OLS to test the assumption of linearity in schooling for our four outcomes. We find significant sheepskin effects in all outcomes and specifications rejecting the linear returns to schooling assumption. Specifically, we run Mincer regressions, then add our dummies for schooling levels and conduct an F-test of the null hypothesis that the coefficients are jointly equal to zero. In Figure A5, we display the pairwise ATEs relative to being a dropout (labeled "Over Dropouts") for each schooling level spaced by the difference in the years of schooling.

Table A16: Years of Schooling Regression: Log Wage

	Log Wage					
Highest Grade Comp.	0.083***	0.037***	0.073***	0.035***	0.040***	0.012
	(0.004)	(0.011)	(0.005)	(0.011)	(0.006)	(0.011)
HS Grad		0.149***		0.145***		0.081*
		(0.041)		(0.043)		(0.042)
GED		0.044		0.011		0.038
		(0.049)		(0.051)		(0.048)
Enroll Coll		0.071**		0.054		0.059*
		(0.034)		(0.034)		(0.033)
Grad. Coll		0.156***		0.142***		0.108**
		(0.044)		(0.045)		(0.043)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.027

Notes: JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A17: Years of Schooling Regression: PV Wage

	PV-Wage					
Highest Grade Comp.	0.122***	0.077***	0.103***	0.076***	0.042***	0.036***
	(0.005)	(0.014)	(0.007)	(0.014)	(0.007)	(0.014)
HS Grad		0.303***		0.260***		0.133**
		(0.052)		(0.055)		(0.052)
GED		-0.019		-0.109*		-0.070
		(0.061)		(0.063)		(0.058)
Enroll Coll		0.017		-0.002		-0.001
		(0.044)		(0.044)		(0.041)
Grad. Coll		0.104*		0.090		0.024
		(0.058)		(0.058)		(0.054)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.000

Notes: JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero .

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Years of Schooling Regression: Health Limits Work

	Health Limits Work					
Highest Grade Comp.	-0.038***	-0.021**	-0.025***	-0.016	-0.014**	-0.010
	(0.004)	(0.010)	(0.005)	(0.010)	(0.006)	(0.011)
HS Grad		-0.109***		-0.068*		-0.052
		(0.036)		(0.039)		(0.040)
GED		0.032		0.069		0.071
		(0.043)		(0.045)		(0.045)
Enroll Coll		-0.016		-0.012		-0.008
		(0.031)		(0.031)		(0.032)
Grad. Coll		-0.027		-0.026		-0.018
		(0.041)		(0.041)		(0.042)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.001	.	0.007

Notes: JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero .

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

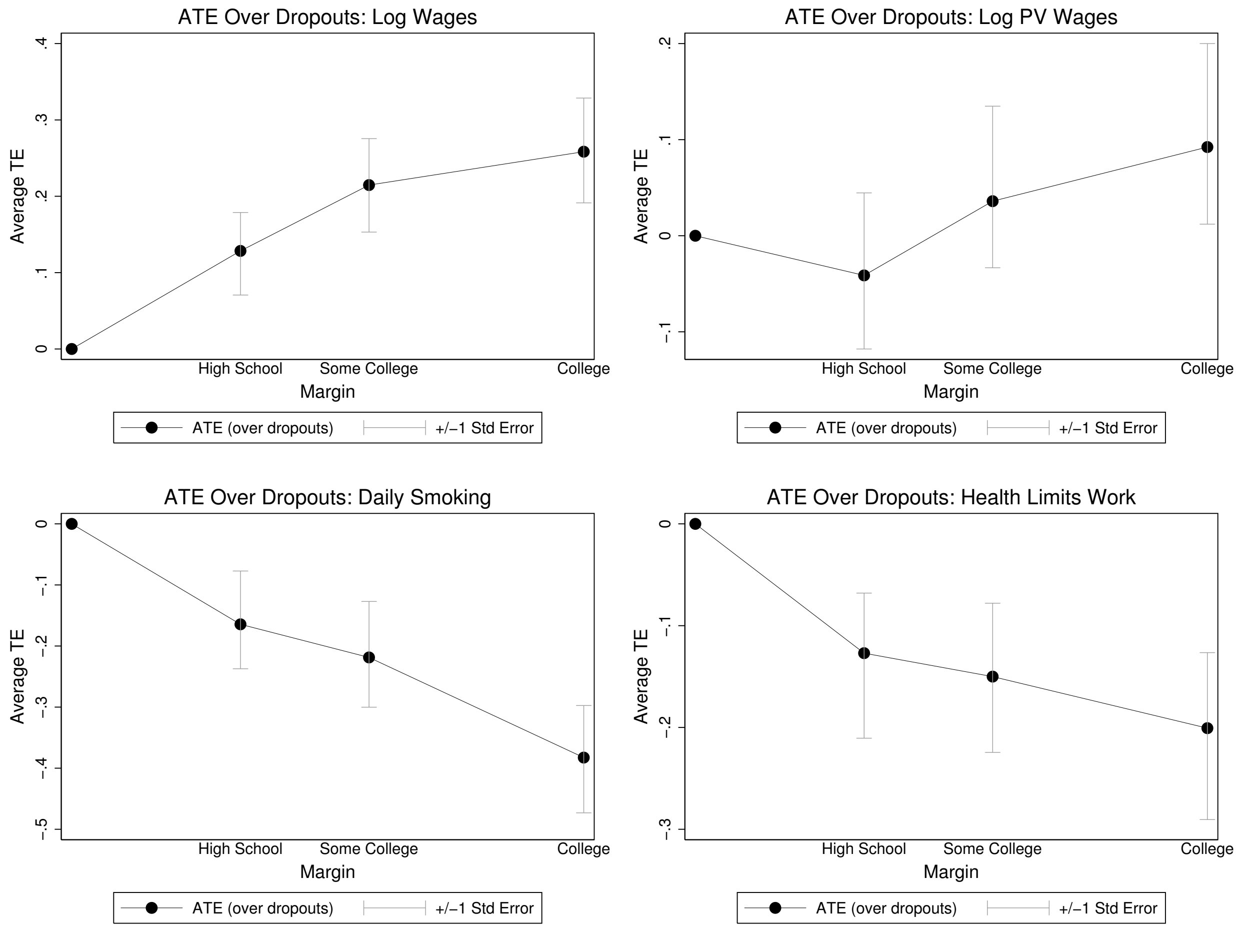
Table A19: Years of Schooling Regression: Smoking

	Smoking					
Highest Grade Comp.	-0.065***	-0.039***	-0.047***	-0.032**	-0.049***	-0.037***
	(0.004)	(0.012)	(0.006)	(0.013)	(0.007)	(0.013)
HS Grad		-0.194***		-0.151***		-0.150***
		(0.045)		(0.048)		(0.049)
GED		0.040		0.042		0.045
		(0.053)		(0.055)		(0.055)
Enroll Coll		0.019		0.022		0.020
		(0.038)		(0.038)		(0.038)
Grad. Coll		-0.076		-0.071		-0.081
		(0.049)		(0.051)		(0.051)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.000

Notes: JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero .

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A5: Plotting Average Treatment Effects by Average Years of Schooling



Notes: Figure displays the estimated pairwise ATE for the full population $(Y_j^k - Y_{j-1}^k)$. The ATEs are spaced out according to the average difference in highest grade completed between each educational group.

A.8.1 Testing the Linearity of the Average Treatment Effect

This section tests if the annualized average treatment effect are linear across schooling decisions. Results are provided in Table A20.[23] Results are reported for both the full-population ATE and the ATE restricted to those who reach the decision node. We can reject linearity for wages for the conditional ATE at the 0.05 level. We can also reject linearity for smoking in the conditional population and linearity for wages in the full population at the 0.10 level, but not the 0.05 level.

[23] See the notes for Table A20 for details on the test's construction.

Table A20: Testing Linearity of the ATE

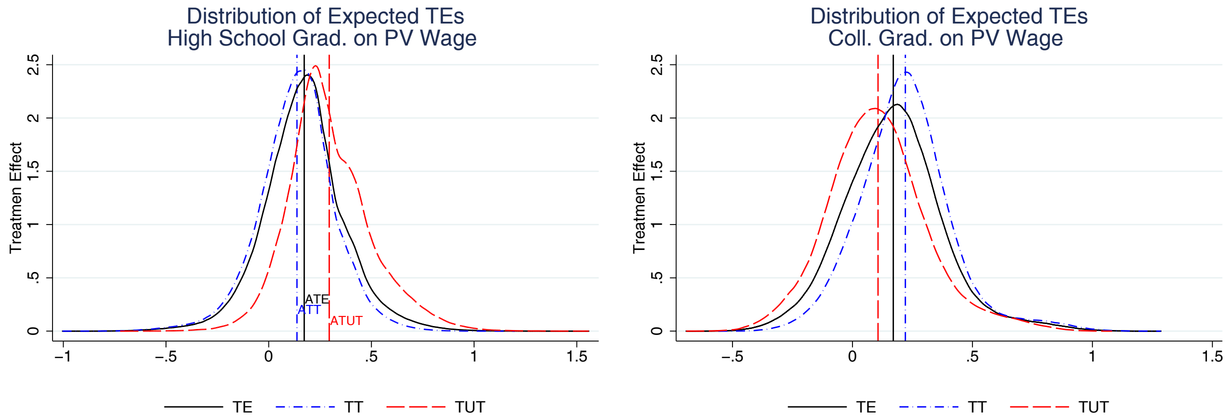
	ATE ($Q_j = 1$)	ATE (full pop)
Wages	0.011	0.090
PV Wages	0.306	0.367
Smoking	0.055	0.653
Health Limits Work	0.353	0.113

Notes: This table reports p-values from a Wald Test for the null hypothesis that the average returns to a year of schooling are linear across schooling decisions. The first column reports the test for the ATE conditional on being at the decision node, while the second column reports the test for the ATE for the full population. Specifically, we test if $\frac{ATE_1}{\bar{q}_1 - \bar{q}_0} - \frac{ATE_2}{\bar{q}_2 - \bar{q}_1} = 0$ and $\frac{ATE_2}{\bar{q}_2 - \bar{q}_1} - \frac{ATE_3}{\bar{q}_3 - \bar{q}_2} = 0$ where the covariance between $\frac{ATE_1}{\bar{q}_1 - \bar{q}_0} - \frac{ATE_2}{\bar{q}_2 - \bar{q}_1}$ and $\frac{ATE_2}{\bar{q}_2 - \bar{q}_1} - \frac{ATE_3}{\bar{q}_3 - \bar{q}_2}$ is estimated using 200 bootstrapped samples. $\bar{q}_j$ is the average years of completed schooling for those at schooling level j.

A.9 Distributions of Treatment Effects

Using the full model it is possible to estimate the distribution of various treatment effects. Figure A6 shows the distribution of expected treatment effects at the choice to graduate from high school and the choice to graduate from college for the log present value of wages. Expectations are computed over the idiosyncratic error terms (ω_s^k). The individual's expected treatment effect is $E_\omega(Y_{s'}^k - Y_s^k) = (\tau_{s'}^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_{s'}^k) - (\tau_s^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k)$, where the variation in the expected treatment effect is coming from the observables ($\boldsymbol{X}$) and the unobserved endowments ($\boldsymbol{\theta}$). The figure shows the distribution of expected treatment effects for everyone at the decision node, the distribution of those that choose to go on ($D_j = 0$), and those that choose not to ($D_j = 1$).

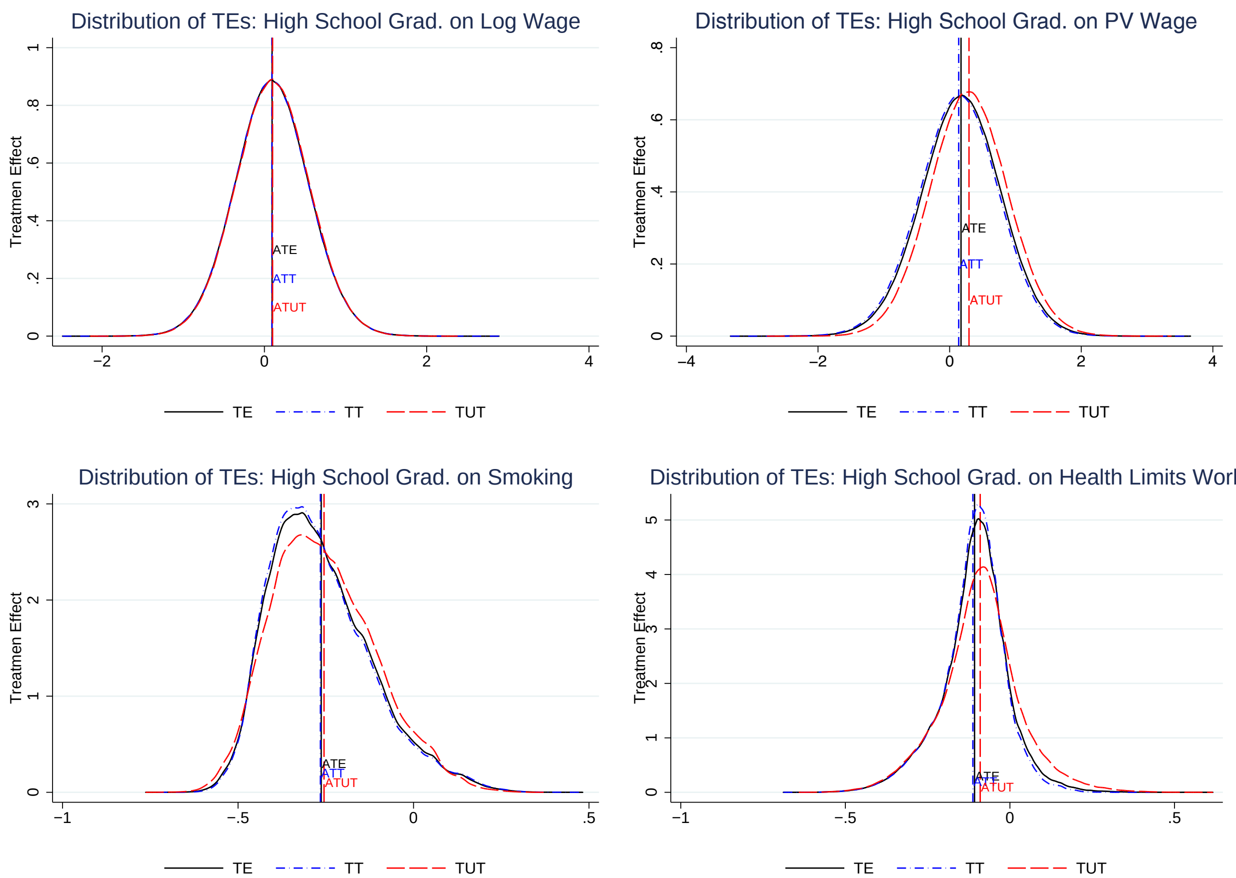
Figure A6: Distributions of Expected Treatment Effects: Log PV Wages



Notes: Distributions of expected treatment effects are for those who reach the educational choice. The expectation is computed over the idiosyncratic error terms (ω_s^k). The individual's expected treatment effect is $E_\omega(Y_{s'}^k - Y_s^k) = \tau_{s'}^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_{s'}^k - (\tau_s^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k)$, where the variation in the expected treatment effect is coming from the observables ($\boldsymbol{X}$) and the unobserved endowments ($\boldsymbol{\theta}$). "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

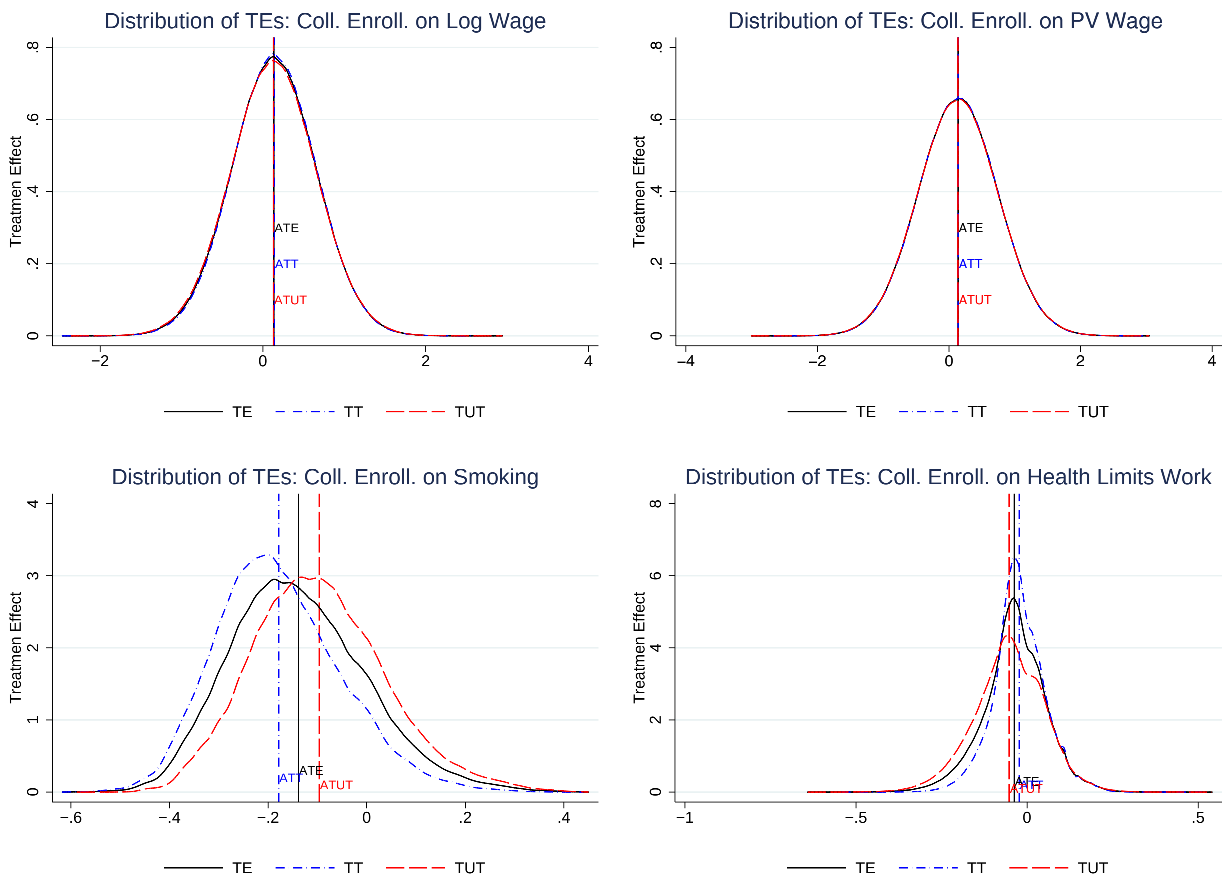
Given the distributions, it is also possible to estimate the percent of individuals who benefit (or are expected to benefit) from a given transition. The model does not impose that individuals make educational choices based on expected gains, making it a testable hypothesis. Examining Figure A6, a portion of each distribution is to the left of 0. This represents the portion of the population that is expected to have lower present value of wages from making the transition. Many individuals do not make the transitions in spite of expected *ex-post* gains, while others make the transitions in spite of expected *ex-post* losses in the present value of wages. We find that the proportion of individuals who benefit is higher for those that choose to graduate from college, while the proportion of individuals who benefit from high school graduation is smaller.

Figure A7: Distributions of Treatment Effects: High School Graduation



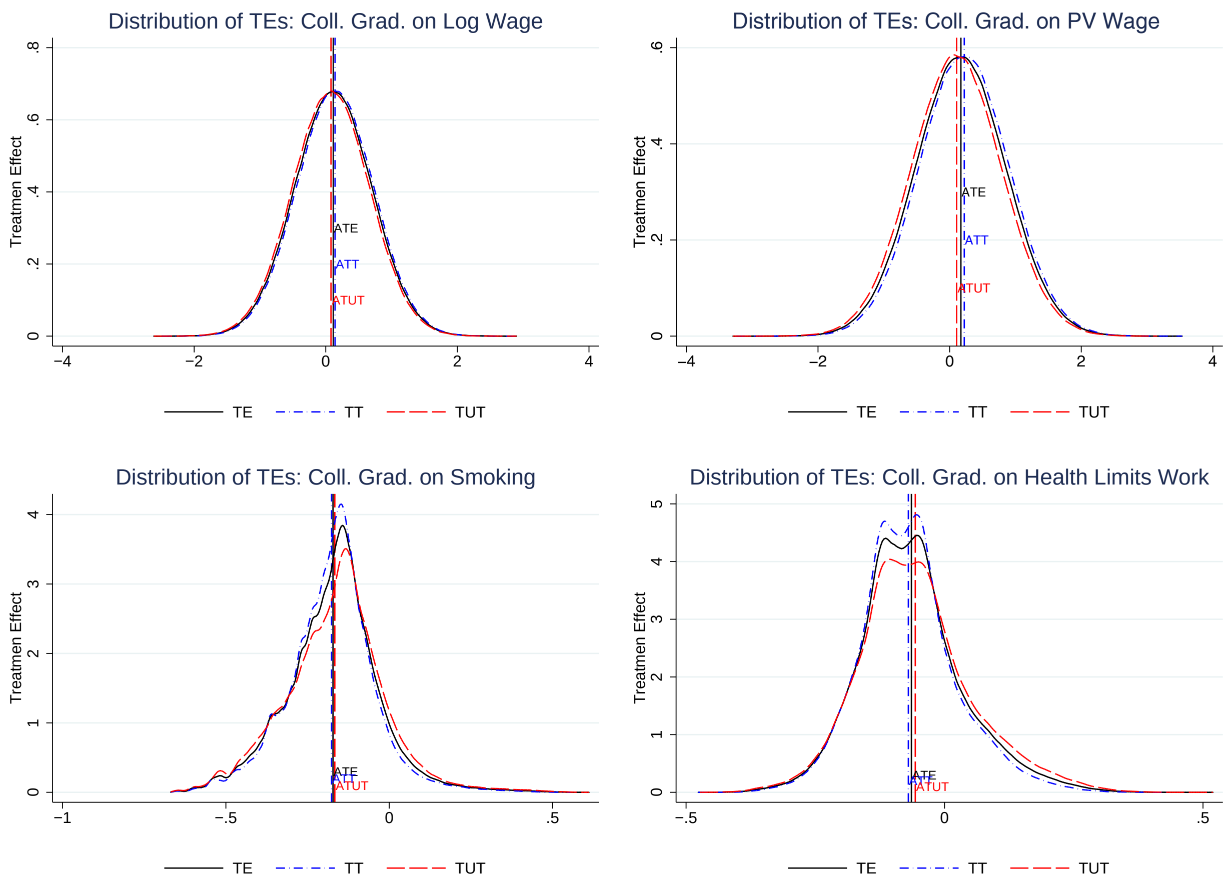
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure A8: Distributions of Treatment Effects: College Enrollment



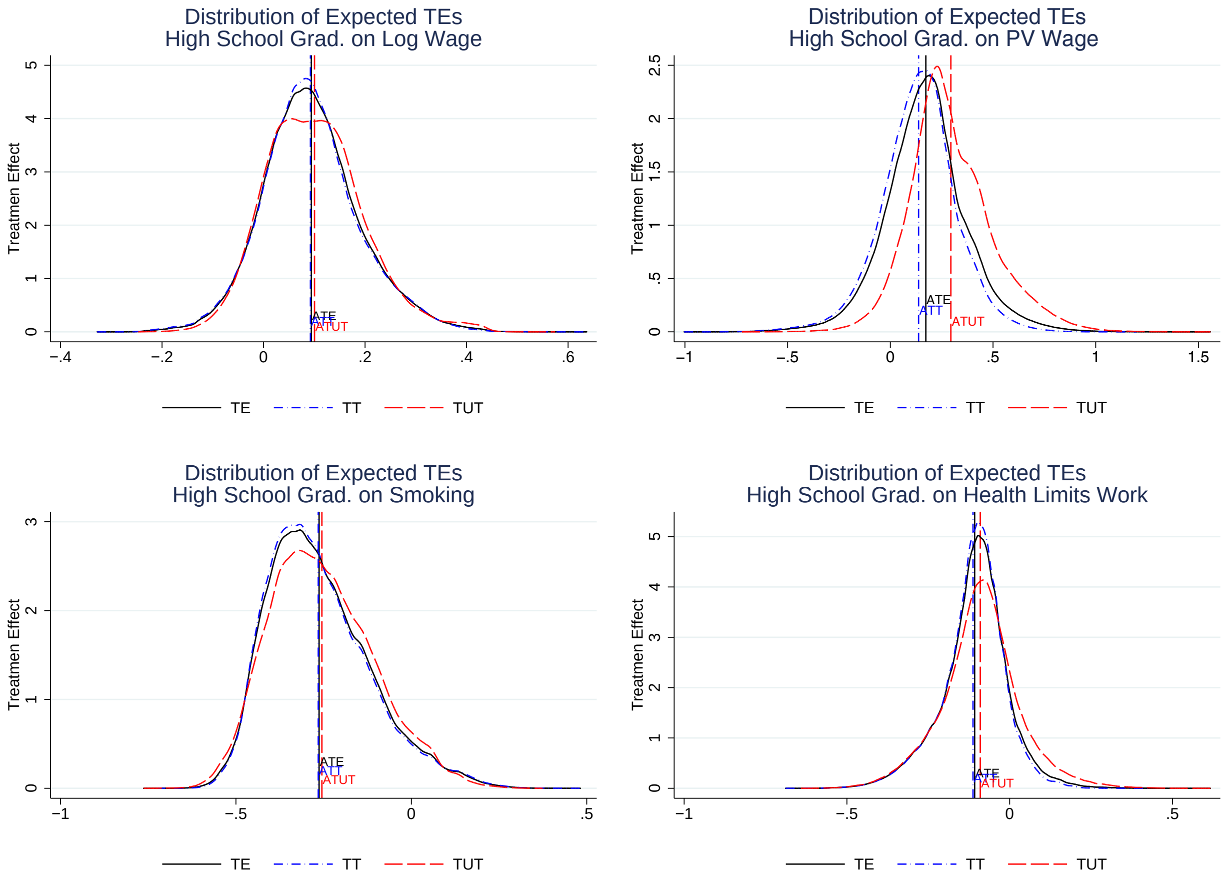
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure A9: Distributions of Treatment Effects: College Graduation



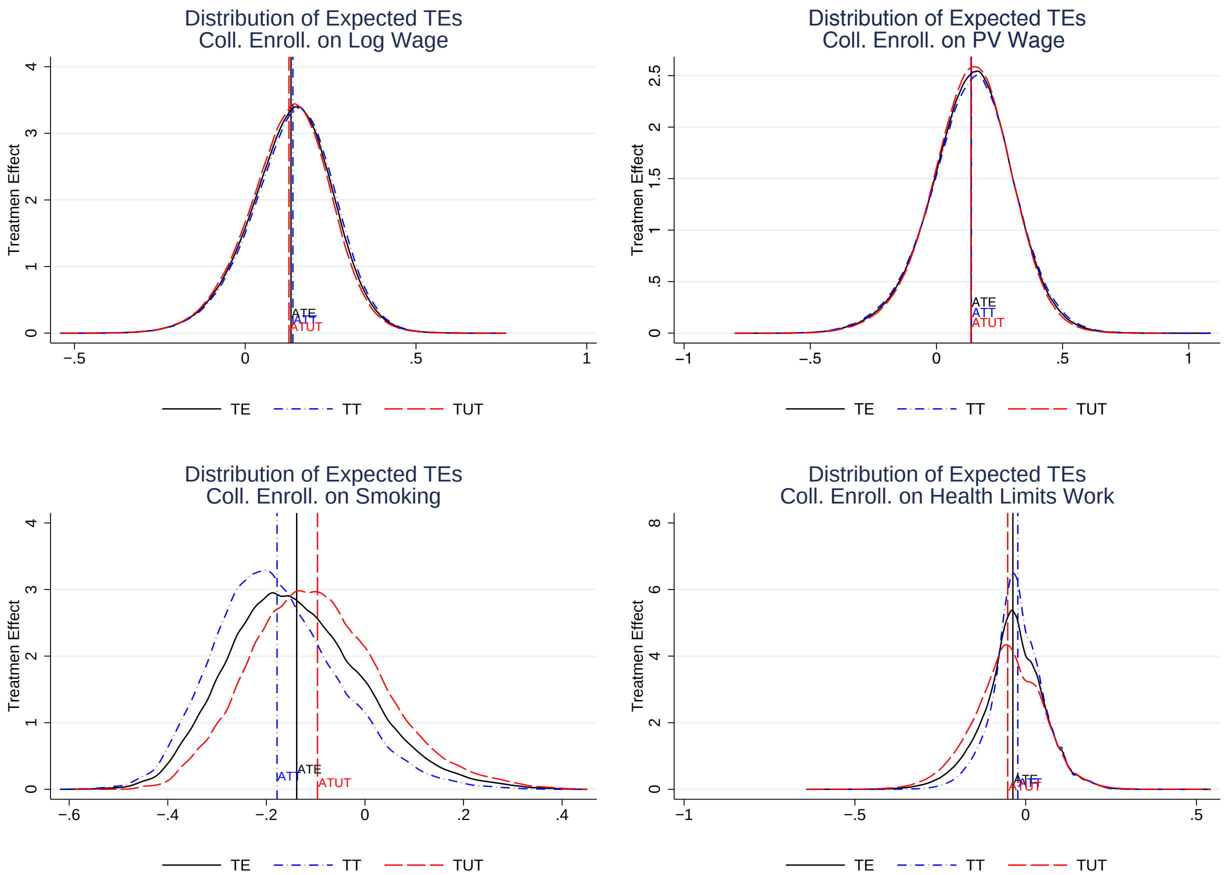
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure A10: Distributions of Expected Treatment Effects: High School Graduation



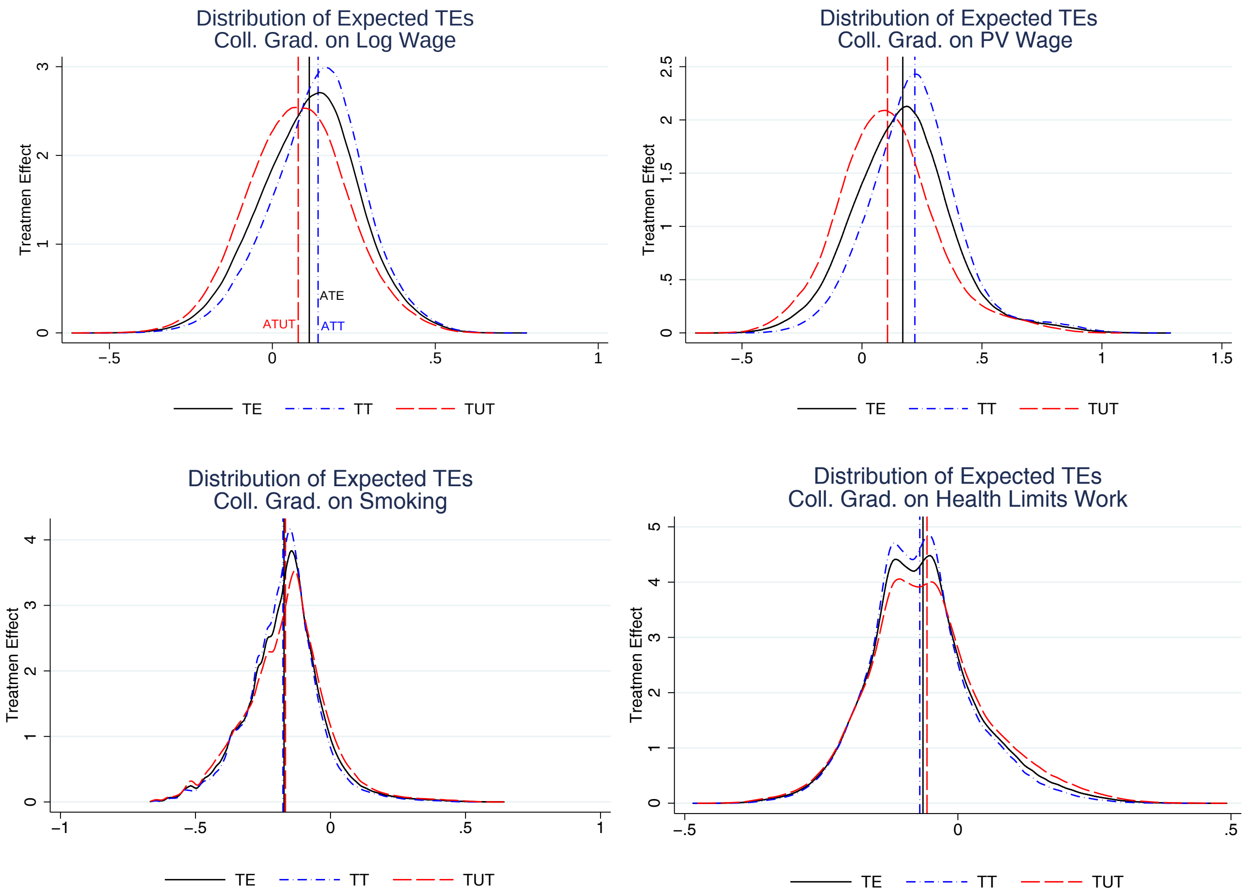
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure A11: Distributions of Expected Treatment Effects: College Enrollment



Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure A12: Distributions of Expected Treatment Effects: College Graduation



Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Table A21: Estimated Percent Who Benefit

Full Population: $Pr(Y_{j+1}^k - Y_j^k > 0)$			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.59	0.53
PV Log Wages	0.62	0.59	0.53
Health Limits Work	0.89	0.68	0.72
Daily Smoking	0.94	0.79	0.88
Conditional on Being at the Decision Node: $Pr(Y_{j+1}^k - Y_j^k > 0 \mid Q_j = 1)$			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.60	0.58
PV Log Wages	0.62	0.59	0.60
Health Limits Work	0.89	0.67	0.78
Daily Smoking	0.94	0.84	0.91
Conditional on Taking the Transition			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.61	0.60
PV Log Wages	0.59	0.59	0.63
Health Limits Work	0.91	0.65	0.80
Daily Smoking	0.94	0.91	0.93
Transition Probabilities: $Pr(D_j = 0 \mid Q_j = 1)$			
	HS Graduation	Enroll in Coll	Grad. College
Prob. of Taking Transition	.775	.514	.558

Notes: Results show the estimated percent who benefit. “Benefit” is defined as reduced probability of smoking, reduced probability of health limiting work, increased wages, or increased PV wages.

Table A22: Spearman Correlations for Counterfactual States Using Simulated Expected Log Wages (age 30)

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.7195	1.0000			
HS Grad.	0.7994	0.8558	1.0000		
Some Coll.	0.8535	0.7338	0.7407	1.0000	
Coll. Grad	0.7077	0.7824	0.7767	0.6986	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table A23: Spearman Correlations for Counterfactual States Using Simulated Expected Log PV Wages

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.8985	1.0000			
HS Grad.	0.9015	0.8327	1.0000		
Some Coll.	0.8321	0.6571	0.7647	1.0000	
Coll. Grad	0.8409	0.8194	0.7275	0.6069	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table A24: Spearman Correlations for Counterfactual States Using Simulated Expected Smoking Age 30

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.6072	1.0000			
HS Grad.	0.4166	0.1254	1.0000		
Some Coll.	0.5503	0.2852	0.1235	1.0000	
Coll. Grad	0.4985	0.4516	0.3892	0.1921	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table A25: Spearman Correlations for Counterfactual States Using Simulated Expected Health Limits Work

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.4076	1.0000			
HS Grad.	0.3104	0.7074	1.0000		
Some Coll.	-0.0823	0.2178	0.1219	1.0000	
Coll. Grad	0.6080	0.6147	0.5162	0.1916	1.0000

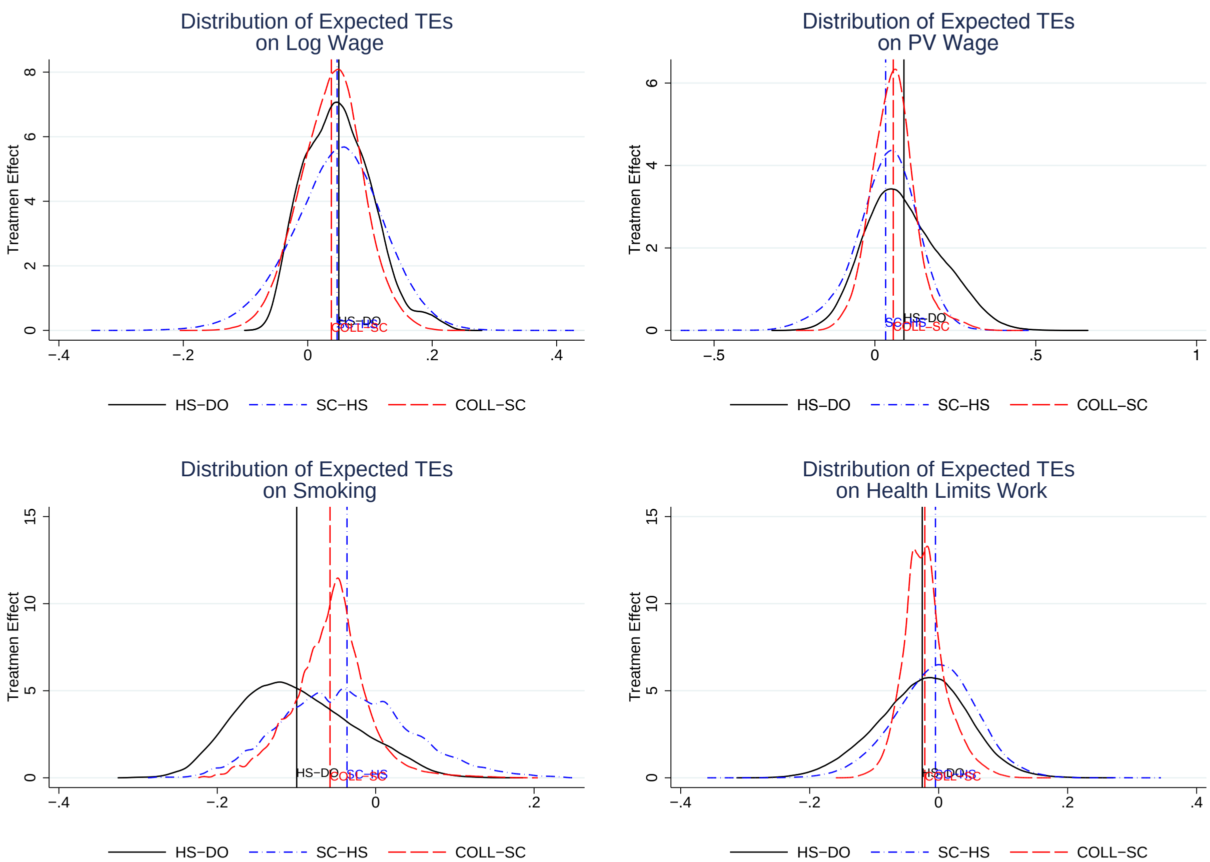
Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

A.10 Distributions of Annualized Returns Across Schooling Transitions

This appendix presents additional information on distributions of annualized returns ($\rho_j = \frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}$) across different schooling levels that supplements the discussion in the text. We present distributions conditional on $Q_j = 1$ (Figure A13). We also present distributions of individual gains inclusive of continuation values (Figure A14).

Figure A13: Annualized Distribution of Returns

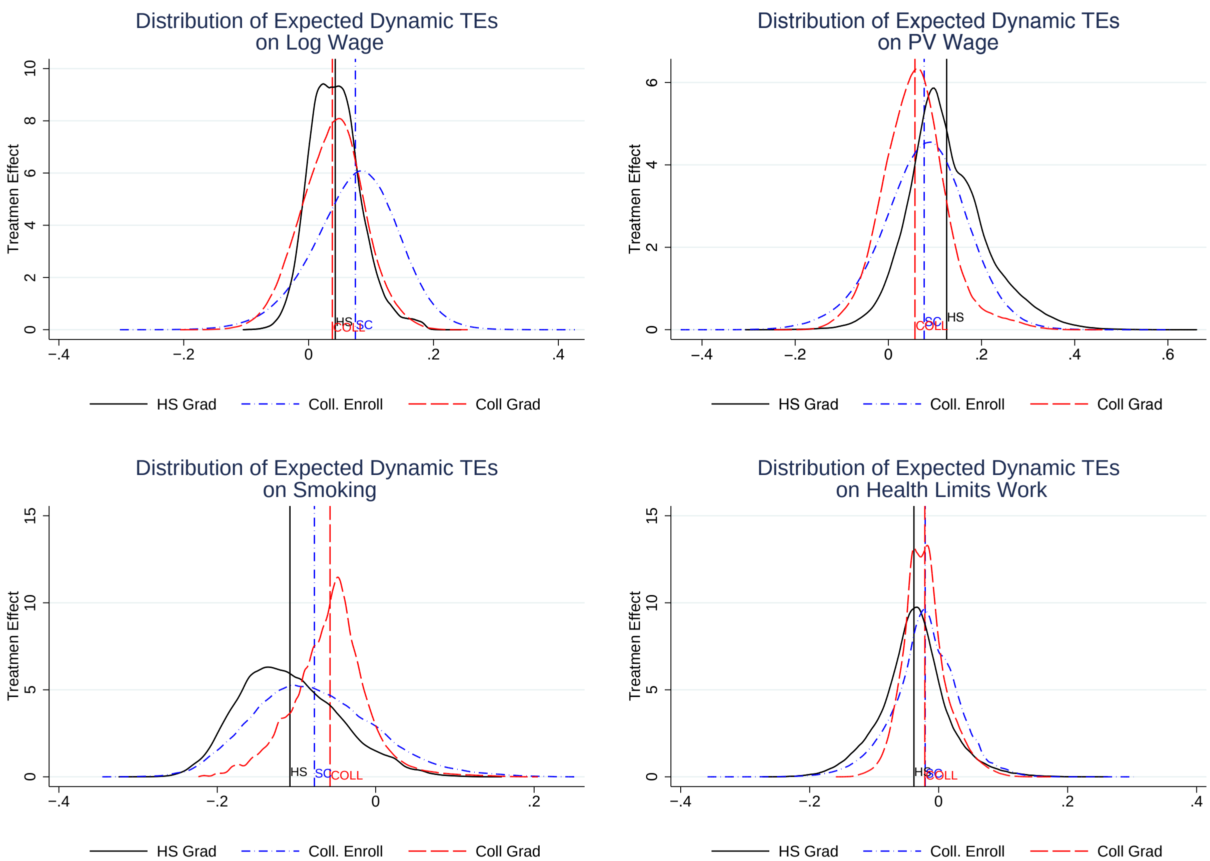
$$\boldsymbol{E_\omega(\rho_j|Q_j=1) = (E_\omega\left[\frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}\right]|Q_{j-1})}$$



Notes: Distributions of treatment effects are for those who reach the educational choice $Q_j = 1$. Distributions have been normalized by the difference in the average number of years of completed schooling between the two educational choices. The returns are $E_\omega(\rho_j|Q_j = 1)$. The vertical lines indicate the means for each distribution.

Distributions of Dynamic Returns This appendix supplements the information that the distributions of annualized returns reported in the text with distributions of annualized dynamic returns $E_\omega\left[\frac{T_j}{q_j - q_{j-1}}\right]$ inclusive of continuation values.

Figure A14: Annualized Distribution of Returns ($E_\omega\left[\frac{T_j}{q_j - q_{j-1}}\right]$)



Notes: Distributions of dynamic treatment effects are for the population who reach the decision. Distributions have been normalized by the difference in the average number of years of completed schooling between the two educational adjacent choices. The shown treatment effects are $E_\omega\left[\frac{T_j}{q_j - q_{j-1}}\right]$, where we are taking the expectation over the idiosyncratic shocks to the outcomes that the agent does not know or act on. The vertical lines indicate the means for each distribution.

A.11 Additional Results From the Simulation of Boosting $\boldsymbol{\theta}$

The observed outcome is given by:

$$Y^k(\boldsymbol{\theta}, \boldsymbol{X}) = \sum_{s \in S} D_s(\boldsymbol{\theta}, \boldsymbol{X}) Y_s^k(\boldsymbol{\theta}, \boldsymbol{X}), \tag{A.5}$$

where both the final educational level $D_s(\boldsymbol{\theta}, \boldsymbol{X})$ and education-specific outcome $Y_s^k(\boldsymbol{\theta}, \boldsymbol{X})$ are affected by changes in $\boldsymbol{\theta}$. For an individual, the effect of increased endowment is given by:

$$Y^k(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}) = \sum_{s \in S} D_s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}) Y_s^k(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}). \tag{A.6}$$

Table A26 shows how the two policy experiments affect final educational choices. The first column shows the proportion of individuals at each level of final education for the lowest decile of endowment. The remaining columns show how this population reallocates to final schooling levels after their endowments change. The increase in both cognitive and non-cognitive endowment cause a number of high school dropouts to increase their final educational attainment. The increase in cognitive ability leads more students to switch from being a dropout to earning a GED, but also has higher college enrollment. An increase in non-cognitive endowment causes somewhat fewer people to change from being a high school dropout, with most switching to high school graduation.

Table A26: Policy Experiment: The Impact of Increasing Endowment in the Bottom Decile on Educational Sorting

Increased Cognitive Endowment						
	Proportion	DO	GED	HS Grad.	Some Coll.	Coll. Grad.
DO	0.372	0.669	0.134	0.162	0.029	0.006
GED	0.107	0.000	0.735	0.195	0.053	0.017
HS	0.401	0.000	0.000	0.867	0.094	0.039
Enroll in Coll.	0.086	0.000	0.000	0.000	0.841	0.159
Grad. Coll.	0.034	0.000	0.000	0.000	0.000	1.000
Increased Socio-Emotional Endowment						
	Proportion	DO	GED	HS Grad.	Some Coll.	Coll. Grad.
DO	0.295	0.754	0.037	0.170	0.031	0.007
GED	0.247	0.000	0.690	0.209	0.072	0.028
HS	0.303	0.000	0.000	0.894	0.073	0.034
Enroll in Coll.	0.101	0.000	0.000	0.000	0.858	0.142
Grad. Coll.	0.054	0.000	0.000	0.000	0.000	1.000

Notes: This table shows the impact of increasing the cognitive (top) or socio-emotional (bottom) endowments of those in the bottom decile of the endowment. All individuals in the bottom decile are given a counterfactual ability which adds the average ability difference between the first and second decile. The first column shows the initial distribution of final schooling of those affected (column sums to 100%). The remaining columns show how what proportion of individuals previously with the row-specific educational choice end up with each new final educational outcome.

A.12 The Robustness of Our Two-Factor Model

A.12.1 Estimating the Number of Factors

As one robustness test, we tried to estimate our model with a third uncorrelated factor that entered only the schooling decisions and the conditional outcomes, but found that this model faced substantial numerical issues and would not converge. Numerical instabilities persisted when trying to estimate the full model jointly for three factors, even when modeling the factors as multivariate normal rather than multivariate mixtures of normals and restricting the factors to be uncorrelated. Given the numerical issues in the three-factor model, this section provides alternative tests for the number of factors that should be used in the model.

Tables A27 and A28 present results from exploratory factor analysis on the full set of cognitive and socio-emotional measures we considered in the NLSY79, including our five measures of risky behavior. The measures include four early risky measures, a violent behavior measure, a reckless behavior measure, six ASVAB tests, and four 9th grade GPAs. All measurements are adjusted for the control variables listed in the table footnotes. As shown in Table A27, principal component factor analysis finds two factors using a scree plot test. As a more robust alternative, we also implement Horn's test (Horn, 1965). As shown in Table A28, Horn's test finds two factors for principal component factor analysis. These exploratory results support the use of two factors.

A.12.2 Comparing our Two-Step Estimation Procedure to Joint Estimation

As described in the text, we estimate our model in two stages. The measurement system and educational choices are estimated in the first stage along with the distribution of latent factors. In the second stage, the school-specific outcomes are estimated, taking the distribution of latent factors as given. Two-stage estimation is computationally simpler, but it also aids in the interpretation of the factors. As we add more outcomes to the measurement system, it

Table A27: Testing the Number of Factors: Results from Eigenvalue or Scree Plot Test

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	5.22807	3.73674	0.4753	0.4753
2	1.49133	0.55935	0.1356	0.6109
3	0.93198	0.08519	0.0847	0.6956
4	0.84678	0.29721	0.0770	0.7726
5	0.54958	0.12020	0.0500	0.8225
6	0.42938	0.06054	0.0390	0.8616
7	0.36884	0.02254	0.0335	0.8951
8	0.34630	0.01958	0.0315	0.9266
9	0.32671	0.05152	0.0297	0.9563
10	0.27519	0.06934	0.0250	0.9813
11	0.20585	.	0.0187	1.0000

Notes: All measures are adjusted for race, parents' education, family income, urban status and region in 1980, age in 1980, and age squared. Results are from an exploratory principal component factor analysis. Criterion: Retain factors with eigenvalues greater than 1. Measures include four early risky measures, a violent behavior measure, a reckless behavior measure, six ASVAB tests, and four 9th grade GPAs.

Table A28: Testing the Number of Factors: Results from Horn's Test

Factor	Adjusted-Eigen	Eigenvalue	
1	4.1101508	5.22807	1.1179192
2	.39898902	1.4913297	1.0923407
3	-.1340403	.93197531	1.0660156
4	-.19610771	.84678427	1.042892
5	-.45441268	.54957654	1.0039892
6	-.55701153	.42938126	.9863928
7	-.60326509	.36883767	.97210276
8	-.62150183	.34629539	.96779722
9	-.61468726	.32671225	.94139951
10	-.64490111	.27518778	.92008889
11	-.68321233	.20584977	.88906211

Notes: All measures are adjusted for race, parents' education, family income, urban status and region in 1980, age in 1980, and age squared. Results are from Horn's parallel analysis on principal component factor analysis. Criterion: Retain factors > 0. Measures include six subtests of ASVAB, GPA in 9th grade in math, language, social studies, and science, and early reckless behavior.

becomes more difficult to interpret the factors. On the other hand, joint estimation allows the unobserved factors to capture additional unobserved correlation between outcomes by construction. In this section, we compare estimates from the model estimated in two stages and the model estimated jointly.

This section provides a number of diagnostics of model-fit under the two-stage and joint estimation procedure. Below, we compare variance decompositions, goodness-of-fit tests, and the estimated model parameters for the model estimated in two stages and the model estimated jointly.

Table A29: Variance Decomposition of Outcomes

Two Step:	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
Log Wages (30)	0.188	0.082	0.010	0.006	0.715
High school dropouts	0.284	0.027	-0.007	0.009	0.688
High school graduates	0.180	0.059	-0.008	0.005	0.764
Some college	0.176	0.004	0.001	0.002	0.817
Four-year college graduate	0.105	0.125	0.014	0.007	0.749
PV Log Wages (30)	0.251	0.091	0.013	0.008	0.637
High school dropouts	0.442	0.162	0.003	0.000	0.393
High school graduates	0.272	0.050	-0.012	0.011	0.678
Some college	0.189	0.011	-0.001	0.000	0.800
Four-year college graduate	0.135	0.083	0.023	0.027	0.731
Smoking Age 30	0.050	0.042	0.027	0.076	0.804
High school dropouts	0.167	0.038	0.023	0.059	0.713
High school graduates	0.068	0.000	0.001	0.007	0.924
Some college	0.099	0.002	0.002	0.009	0.888
Four-year college graduate	0.097	0.021	0.015	0.046	0.822
Health Limits Work	0.060	0.046	0.014	0.020	0.860
High school dropouts	0.116	0.037	-0.018	0.039	0.826
High school graduates	0.076	0.038	0.008	0.008	0.870
Some college	0.071	0.000	0.000	0.008	0.921
Four-year college graduate	0.125	0.045	0.007	0.005	0.817
Joint:	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
Log Wages (30)					
High school dropouts	0.290	0.046	-0.020	0.035	0.650
High school graduates	0.178	0.081	-0.019	0.017	0.744
Some college	0.177	0.004	0.002	0.003	0.813
Four-year college graduate	0.108	0.147	0.025	0.016	0.703
PV Log Wages (30)					
High school dropouts	0.438	0.184	-0.021	0.009	0.390
High school graduates	0.270	0.071	-0.021	0.024	0.656
Some college	0.193	0.012	0.001	0.000	0.794
Four-year college graduate	0.138	0.113	0.032	0.035	0.681
Smoking Age 30					
High school dropouts	0.166	0.040	0.028	0.074	0.692
High school graduates	0.069	0.001	0.001	0.004	0.926
Some college	0.099	0.001	0.002	0.014	0.884
Four-year college graduate	0.092	0.026	0.019	0.053	0.809
Health Limits Work					
High school dropouts	0.114	0.055	-0.030	0.064	0.797
High school graduates	0.074	0.047	0.006	0.003	0.870
Some college	0.073	0.000	-0.000	0.012	0.915
Four-year college graduate	0.123	0.056	0.014	0.013	0.794

Notes: Columns show the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socio-emotional 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 characteristics, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\mathbf{X'}\boldsymbol{\beta}) + var(\boldsymbol{\theta'}\boldsymbol{\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\mathbf{X'}\boldsymbol{\beta}) + var(\boldsymbol{\theta'}\boldsymbol{\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha'}\boldsymbol{\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE}) + var(\theta_{SE}\alpha_{SE})$. In the legend above, for continuous outcomes, "Observables" is $var(\mathbf{X'}\boldsymbol{\beta})/var(Y)$, "Cognitive" is $var(\theta_C\alpha_C)/var(Y)$, "Covariance" is $2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE})/var(Y)$, and "Socio-Emotional" is $var(\theta_{SE}\alpha_{SE})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$.

Table A30: Goodness of Fit - Schooling Choice (two-step and joint estimation)

Two-Stage:	Data	Model	p-value
High School Dropout	0.131	0.122	0.980
High School Graduate	0.370	0.377	0.989
Some College	0.168	0.176	0.982
College Graduate	0.230	0.222	0.986
Joint:	Data	Model	p-value
High School Dropout	0.131	0.123	0.981
High School Graduate	0.370	0.376	0.989
Some College	0.168	0.176	0.982
College Graduate	0.230	0.222	0.986

Notes: The simulated data (Model) contains one million observations generated from the model estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is *Model=Data*.

Table A31: Goodness of Fit - Discrete Outcomes (two-step and joint estimation)

Two-Step:	Actual	Model	p-value[a]
Smoking Age 30	0.385	0.387	0.997
High school dropouts	0.674	0.650	0.959
High school graduates	0.390	0.383	0.989
Some college	0.337	0.339	0.995
Four-year college graduate	0.146	0.166	0.955
Health Limits Work	0.227	0.226	0.997
High school dropouts	0.392	0.412	0.968
High school graduates	0.232	0.229	0.994
Some college	0.184	0.179	0.992
Four-year college graduate	0.091	0.099	0.980
Joint:	Actual	Model	p-value[a]
Smoking Age 30	0.385	0.432	0.922
High school dropouts	0.674	0.653	0.964
High school graduates	0.390	0.384	0.990
Some college	0.337	0.338	0.997
Four-year college graduate	0.146	0.168	0.950
Health Limits Work	0.227	0.274	0.912
High school dropouts	0.392	0.410	0.972
High school graduates	0.232	0.231	0.999
Some college	0.184	0.176	0.985
Four-year college graduate	0.091	0.098	0.983

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is that the model predictions fits the data.

Table A32: Goodness of Fit - Continuous Outcomes (two-step and joint estimation)

Two-Stage	Mean		Std Dev		
	Actual	Model	Actual	Model	p-value
Log Wages (30)	2.612	2.604	0.229	0.223	0.132
High school dropouts	2.291	2.247	0.135	0.130	0.000
High school graduates	2.531	2.528	0.184	0.182	0.637
Some college	2.665	2.677	0.207	0.200	0.283
Four-year college graduate	2.932	2.949	0.188	0.186	0.039
PV Log Wages (30)	12.315	12.317	0.397	0.395	0.876
High school dropouts	11.787	11.681	0.366	0.391	0.000
High school graduates	12.275	12.275	0.273	0.262	0.983
Some college	12.422	12.432	0.257	0.255	0.499
Four-year college graduate	12.764	12.817	0.266	0.272	0.000
Joint:	Mean		Std Dev		
	Actual	Model	Actual	Model	p-value
Log Wages (30)	2.612	2.474	0.229	0.205	0.000
High school dropouts	2.291	2.241	0.135	0.132	0.000
High school graduates	2.531	2.527	0.184	0.182	0.503
Some college	2.665	2.682	0.207	0.201	0.134
Four-year college graduate	2.932	2.947	0.188	0.188	0.072
PV Log Wages (30)	12.315	12.090	0.397	0.387	0.000
High school dropouts	11.787	11.678	0.366	0.388	0.000
High school graduates	12.275	12.274	0.273	0.262	0.952
Some college	12.422	12.434	0.257	0.255	0.396
Four-year college graduate	12.764	12.811	0.266	0.276	0.000

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) The p-values reported are from a T-test for the equivalence of the means where the null hypothesis is that the model predictions fits the data.

Table A33: Estimates for Log Wages at Age 30 (comparing joint and two-stage estimation)

Two-Stage Estimation:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.200	0.068	-0.278	0.089	-0.229	0.052	-0.247	0.081	-0.066	0.079
Hispanic	-0.178	0.077	-0.050	0.143	-0.004	0.065	-0.136	0.103	-0.009	0.109
Broken Home	-0.022	0.047	0.139	0.073	0.064	0.038	-0.082	0.059	-0.063	0.056
Number of Sibs	-0.006	0.009	0.009	0.014	-0.007	0.007	0.008	0.011	0.006	0.010
Mother's HGC	0.004	0.011	0.006	0.017	-0.001	0.008	0.014	0.012	0.013	0.009
Father's HGC	0.008	0.009	0.019	0.012	0.012	0.006	-0.006	0.009	0.006	0.007
Age	0.151	0.214	0.331	0.329	-0.136	0.133	0.191	0.211	0.097	0.175
Age Squared	-0.004	0.006	-0.007	0.009	0.004	0.003	-0.005	0.005	-0.002	0.005
Fam. Income 1979	0.008	0.003	0.008	0.004	0.007	0.002	0.008	0.002	0.005	0.001
Constant	0.591	2.015	-1.690	3.079	3.251	1.255	0.451	1.982	1.361	1.650
Local Unemp.	-1.021	1.192	0.905	1.605	0.511	0.759	-1.520	1.199	-0.984	1.050
Northeast 30	0.291	0.073	0.125	0.115	0.070	0.041	0.128	0.070	0.183	0.050
South 30	0.078	0.058	0.024	0.089	-0.046	0.038	0.016	0.063	-0.025	0.049
West 30	0.060	0.076	0.005	0.107	0.046	0.045	0.062	0.066	-0.001	0.056
Urban 30	0.055	0.054	0.108	0.084	0.098	0.033	0.117	0.060	0.139	0.056
Factor 1	0.095	0.052	0.147	0.059	0.157	0.024	0.040	0.042	0.233	0.042
Factor 2	-0.056	0.054	0.076	0.073	-0.049	0.029	0.032	0.051	0.057	0.045

Joint Estimation:

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
Black	-0.208	0.067	-0.273	0.089	-0.234	0.053	-0.246	0.081	-0.066	0.080
Hispanic	-0.176	0.077	-0.034	0.142	-0.005	0.065	-0.133	0.103	-0.000	0.110
Broken Home	-0.020	0.047	0.135	0.073	0.062	0.038	-0.082	0.059	-0.067	0.057
Number of Sibs	-0.005	0.009	0.009	0.014	-0.007	0.007	0.007	0.011	0.006	0.010
Mother's HGC	0.004	0.011	0.009	0.017	-0.001	0.008	0.015	0.012	0.014	0.010
Father's HGC	0.009	0.009	0.019	0.012	0.011	0.006	-0.005	0.009	0.008	0.007
Age	0.157	0.212	0.334	0.328	-0.122	0.134	0.189	0.211	0.079	0.178
Age Squared	-0.004	0.005	-0.007	0.009	0.004	0.003	-0.005	0.005	-0.002	0.005
Fam. Income 1979	0.008	0.003	0.008	0.004	0.007	0.002	0.008	0.002	0.005	0.001
Constant	0.506	2.002	-1.754	3.065	3.132	1.265	0.449	1.982	1.450	1.680
Local Unemp	-1.036	1.173	0.900	1.592	0.524	0.751	-1.545	1.198	-1.022	1.041
Northeast 30	0.284	0.073	0.130	0.115	0.068	0.041	0.131	0.070	0.188	0.050
South 30	0.072	0.058	0.023	0.088	-0.044	0.038	0.017	0.063	-0.022	0.049
West 30	0.052	0.075	-0.002	0.106	0.047	0.045	0.063	0.066	0.003	0.056
Urban 30	0.057	0.053	0.106	0.083	0.098	0.033	0.117	0.060	0.139	0.056
Factor 1	0.123	0.053	0.175	0.060	0.183	0.025	0.044	0.042	0.256	0.046
Factor 2	-0.112	0.055	0.089	0.078	-0.088	0.031	0.041	0.054	0.089	0.050

Notes: The top panel shows results where the model is estimated in two stages. In the first stage the measurement system and educational choice are jointly estimated with the distribution of latent factors. In the second stage, the schooling-specific outcomes are estimated taking the distribution of latent factors from the first stage as given. In the bottom panel, the measurement system, educational choice, and schooling-specific outcomes are estimated jointly with the estimation of the distribution of the factors. In the second stage, outcomes directly inform the factors.

Table A34: Estimates for Log Present Value of Wages (comparing joint and two-stage estimation)

Two-Stage Estimation:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.745	0.089	-0.351	0.114	-0.345	0.060	-0.318	0.091	-0.104	0.101
Hispanic	-0.152	0.113	-0.076	0.192	0.200	0.080	-0.315	0.109	0.315	0.137
Broken Home	-0.079	0.065	-0.044	0.095	-0.042	0.044	-0.122	0.066	-0.120	0.070
Number of Sibs	-0.037	0.012	0.031	0.018	-0.005	0.008	-0.003	0.014	-0.002	0.013
Mother's HGC	0.026	0.015	0.032	0.022	0.018	0.010	0.015	0.014	0.036	0.012
Father's HGC	0.016	0.012	0.021	0.016	0.019	0.007	-0.010	0.010	0.003	0.009
Age	-0.270	0.267	-0.300	0.379	0.012	0.143	0.025	0.221	-0.229	0.195
Age Squared	0.007	0.007	0.008	0.010	-0.000	0.004	-0.001	0.006	0.006	0.005
Fam. Income 1979	0.018	0.004	0.017	0.005	0.012	0.002	0.008	0.002	0.008	0.002
Constant	14.028	2.571	13.716	3.621	11.454	1.382	11.966	2.111	14.092	1.879
Urban 17	0.216	0.070	0.038	0.111	0.045	0.039	0.175	0.062	0.066	0.060
South 17	0.066	0.076	0.157	0.115	0.024	0.045	0.029	0.069	0.031	0.058
West 17	-0.144	0.101	-0.157	0.137	-0.027	0.052	0.116	0.076	-0.143	0.078
Northeast 17	0.162	0.097	-0.001	0.147	0.035	0.048	0.101	0.076	0.158	0.058
Factor 1	0.429	0.072	0.408	0.075	0.180	0.029	0.079	0.046	0.230	0.053
Factor 2	0.020	0.076	0.017	0.093	-0.089	0.035	-0.010	0.056	0.138	0.057

Joint Estimation:

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.728	0.091	-0.344	0.115	-0.348	0.060	-0.318	0.091	-0.104	0.101
Hispanic	-0.144	0.115	-0.052	0.192	0.197	0.080	-0.310	0.109	0.333	0.138
Broken Home	-0.066	0.066	-0.048	0.096	-0.045	0.044	-0.124	0.066	-0.127	0.071
Number of Sibs	-0.035	0.013	0.029	0.018	-0.006	0.008	-0.003	0.014	-0.003	0.013
Mother's HGC	0.024	0.015	0.034	0.022	0.018	0.010	0.016	0.014	0.036	0.012
Father's HGC	0.017	0.013	0.022	0.016	0.019	0.007	-0.009	0.010	0.006	0.009
Age	-0.240	0.276	-0.301	0.381	0.025	0.144	0.014	0.221	-0.249	0.198
Age Squared	0.007	0.007	0.009	0.010	-0.000	0.004	-0.001	0.006	0.006	0.005
Fam. Income 1979	0.018	0.004	0.017	0.005	0.012	0.002	0.008	0.002	0.008	0.002
Constant	13.688	2.652	13.682	3.645	11.341	1.394	12.038	2.112	14.183	1.908
Urban 17	0.206	0.071	0.040	0.111	0.047	0.039	0.176	0.062	0.069	0.060
South 17	0.044	0.077	0.158	0.115	0.026	0.045	0.033	0.069	0.037	0.058
West 17	-0.144	0.102	-0.152	0.136	-0.025	0.052	0.119	0.077	-0.134	0.078
Northeast 17	0.140	0.098	0.024	0.147	0.032	0.048	0.109	0.076	0.165	0.058
Factor 1	0.444	0.073	0.440	0.076	0.212	0.030	0.082	0.046	0.273	0.057
Factor 2	-0.103	0.081	0.071	0.100	-0.130	0.037	0.012	0.059	0.159	0.064

Notes: The top panel shows results where the model is estimated in two stages. In the first stage, the measurement system and educational choice are jointly estimated with the distribution of latent factors. In the second stage, the schooling-specific outcomes are estimated taking the distribution of latent factors from the first stage as given. In the bottom panel, the measurement system, educational choice, and schooling-specific outcomes are estimated jointly with the estimation of the distribution of the factors. In the second stage, outcomes directly inform the factors.

Table A35: Estimates for Smoking at Age 30 (comparing joint and two-stage estimation)

Two-Stage Estimation:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.199	0.277	-0.106	0.271	0.216	0.180	-0.349	0.288	-0.081	0.359
Hispanic	-0.943	0.353	-1.298	0.467	-0.296	0.245	-0.413	0.342	0.520	0.431
Broken Home	0.317	0.209	-0.225	0.222	0.037	0.129	0.623	0.199	-0.008	0.253
Number of Sibs	0.104	0.043	-0.020	0.045	-0.003	0.025	0.053	0.038	-0.019	0.043
Mother's HGC	0.004	0.046	-0.072	0.054	0.018	0.028	0.031	0.042	-0.076	0.040
Father's HGC	0.054	0.040	0.031	0.038	0.046	0.022	0.008	0.029	0.076	0.032
Age	0.240	0.822	0.393	0.910	-0.343	0.417	0.543	0.675	-0.297	0.685
Age Squared	-0.004	0.021	-0.008	0.024	0.009	0.011	-0.014	0.017	0.009	0.018
Fam. Income 1979	0.013	0.013	-0.006	0.012	-0.009	0.005	0.004	0.007	-0.003	0.005
Constant	-4.229	7.854	-3.494	8.633	2.423	4.018	-6.544	6.445	1.353	6.608
Northeast 30	0.073	0.327	-0.323	0.361	0.132	0.139	-0.281	0.239	-0.199	0.215
South 30	0.023	0.246	-0.432	0.281	0.119	0.129	-0.120	0.208	0.010	0.198
West 30	-0.291	0.303	-0.062	0.330	-0.212	0.155	0.018	0.217	-0.425	0.267
Urban 30	-0.104	0.225	0.147	0.257	0.043	0.113	0.112	0.196	0.421	0.286
Factor 1	-0.341	0.229	-0.385	0.183	-0.023	0.083	-0.068	0.135	-0.233	0.181
Factor 2	-0.444	0.233	-0.166	0.215	-0.131	0.103	-0.153	0.164	-0.361	0.195

Joint Estimation:

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
Black	-0.200	0.278	-0.110	0.272	0.221	0.180	-0.346	0.289	-0.085	0.361
Hispanic	-0.955	0.357	-1.312	0.470	-0.296	0.245	-0.419	0.342	0.504	0.433
Broken Home	0.333	0.212	-0.227	0.223	0.036	0.129	0.625	0.199	-0.006	0.254
Number of Sibs	0.106	0.043	-0.019	0.046	-0.003	0.025	0.054	0.038	-0.018	0.043
Mother's HGC	0.002	0.046	-0.072	0.054	0.019	0.028	0.029	0.042	-0.076	0.040
Father's HGC	0.056	0.040	0.030	0.038	0.046	0.022	0.007	0.029	0.073	0.033
Age	0.211	0.830	0.382	0.916	-0.345	0.417	0.544	0.676	-0.284	0.690
Age Squared	-0.003	0.022	-0.007	0.024	0.009	0.011	-0.014	0.017	0.008	0.018
Fam. Income 1979	0.013	0.014	-0.006	0.012	-0.009	0.005	0.004	0.007	-0.003	0.005
Constant	-3.985	7.936	-3.361	8.682	2.438	4.015	-6.513	6.458	1.305	6.663
Northeast 30	0.063	0.330	-0.333	0.361	0.134	0.139	-0.289	0.240	-0.213	0.217
South 30	0.026	0.248	-0.431	0.281	0.118	0.129	-0.126	0.208	0.003	0.199
West 30	-0.301	0.305	-0.067	0.331	-0.213	0.155	0.019	0.218	-0.432	0.268
Urban 30	-0.112	0.227	0.146	0.258	0.042	0.113	0.109	0.197	0.422	0.287
Factor 1	-0.352	0.230	-0.391	0.186	-0.040	0.084	-0.055	0.136	-0.264	0.186
Factor 2	-0.500	0.228	-0.194	0.215	-0.097	0.103	-0.192	0.169	-0.393	0.201

Notes: The top panel shows results where the model is estimated in two stages. In the first stage, the measurement system and educational choice are jointly estimated with the distribution of latent factors. In the second stage, the schooling-specific outcomes are estimated taking the distribution of latent factors from the first stage as given. In the bottom panel, the measurement system, educational choice, and schooling-specific outcomes are estimated jointly with the estimation of the distribution of the factors. In the second stage, outcomes directly inform the factors.

Table A36: Outcome Model: Estimates for Health Limits Work (comparing joint and two-stage estimation)

Two-Stage Estimation:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	0.360	0.228	0.033	0.242	0.102	0.174	0.173	0.269	0.142	0.321
Hispanic	-0.178	0.292	-0.352	0.391	0.129	0.214	0.204	0.329	-0.782	0.612
Broken Home	0.079	0.173	-0.061	0.199	-0.051	0.132	0.303	0.195	0.250	0.238
Number of Sibs	0.002	0.033	-0.032	0.039	-0.001	0.025	0.038	0.038	0.019	0.044
Mother's HGC	-0.009	0.040	-0.066	0.048	-0.016	0.027	-0.021	0.042	-0.015	0.043
Father's HGC	0.011	0.033	-0.033	0.034	-0.051	0.021	0.023	0.030	-0.003	0.031
Age	0.034	0.722	-0.388	0.803	0.869	0.436	0.205	0.675	-0.349	0.725
Age Squared	0.002	0.019	0.012	0.021	-0.021	0.011	-0.006	0.017	0.012	0.018
Fam. Income 1979	-0.012	0.011	-0.013	0.010	-0.000	0.006	0.003	0.007	0.004	0.005
Constant	-1.774	6.969	4.121	7.706	-8.930	4.214	-2.751	6.442	0.929	7.047
Urban 17	-0.165	0.191	-0.002	0.237	0.204	0.120	-0.280	0.187	-0.113	0.224
South 17	0.137	0.206	-0.288	0.242	-0.054	0.136	-0.195	0.207	0.160	0.216
West 17	0.407	0.277	0.060	0.275	0.249	0.153	0.069	0.224	0.407	0.270
Northeast 17	0.374	0.259	-0.219	0.305	0.034	0.140	-0.418	0.242	0.067	0.224
Factor 1	-0.312	0.190	-0.269	0.159	-0.309	0.085	0.013	0.140	-0.347	0.175
Factor 2	0.337	0.201	-0.169	0.195	-0.146	0.100	-0.147	0.167	-0.124	0.203

Joint Estimation:

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
Black	0.362	0.231	0.028	0.243	0.105	0.175	0.171	0.269	0.143	0.323
Hispanic	-0.205	0.300	-0.368	0.394	0.129	0.216	0.195	0.329	-0.824	0.627
Broken Home	0.079	0.176	-0.061	0.200	-0.046	0.133	0.306	0.195	0.265	0.241
Number of Sibs	0.001	0.034	-0.031	0.039	0.001	0.025	0.039	0.038	0.020	0.045
Mother's HGC	-0.008	0.041	-0.066	0.048	-0.016	0.028	-0.022	0.042	-0.016	0.043
Father's HGC	0.008	0.033	-0.034	0.034	-0.050	0.021	0.022	0.030	-0.008	0.031
Age	0.008	0.737	-0.392	0.806	0.849	0.439	0.208	0.675	-0.297	0.733
Age Squared	0.003	0.019	0.012	0.021	-0.021	0.011	-0.006	0.017	0.011	0.019
Fam. Income 1979	-0.012	0.011	-0.013	0.010	0.000	0.006	0.003	0.007	0.004	0.005
Constant	-1.537	7.107	4.178	7.734	-8.752	4.243	-2.747	6.450	0.575	7.128
Urban 17	-0.154	0.195	-0.005	0.237	0.200	0.121	-0.281	0.188	-0.119	0.225
South 17	0.158	0.209	-0.290	0.243	-0.057	0.137	-0.200	0.208	0.145	0.217
West 17	0.408	0.281	0.057	0.275	0.246	0.154	0.070	0.225	0.396	0.272
Northeast 17	0.402	0.263	-0.228	0.307	0.036	0.141	-0.428	0.243	0.054	0.226
Factor 1	-0.384	0.194	-0.281	0.161	-0.340	0.087	0.004	0.141	-0.390	0.182
Factor 2	0.434	0.203	-0.181	0.196	-0.093	0.100	-0.173	0.172	-0.196	0.210

Notes: The top panel shows results where the model is estimated in two stages. In the first stage, the measurement system and educational choice are jointly estimated with the distribution of latent factors. In the second stage, the schooling-specific outcomes are estimated taking the distribution of latent factors from the first stage as given. In the bottom panel, the measurement system, educational choice, and schooling-specific outcomes are estimated jointly with the estimation of the distribution of the factors. In the second stage, outcomes directly inform the factors.

A.12.3 Testing How Well Correlations from the Simulation Match Correlations in the Data

One question is how well our model captures forward-looking behavior of the agent. One simple test is to see how well our model reproduces the correlations found in the data. Specifically, we can calculate $COV(Y^k, D_j)$ for all outcomes k and choices j where Y^k is the observed outcome for each individual. This correlation can be estimated in both the simulation and the data, and we can test against the null hypothesis that they are equal.

Table A37 shows the p-values from the test of the null hypothesis of if $COV(Y^k, D_j)$ are equal in the simulation and the data. We fail to reject the null of equality for all outcome-decision pairs, except for the choice to enroll in college with log present value of wages and the choice to enroll in college with being a regular smoker at age 30. The p-values are very large for wages, showing that we fit the covariance found in the data with the model quite well.

Table A37: Test of Equality of Sample Covariance and Simulated Covariance $\boldsymbol{COV(Y^k, D_j)}$

	HS	SC	Coll
Wages	0.959	0.688	0.990
PV Wages	0.571	0.004	0.160
Smoking	0.841	0.059	0.877
Work-Limited	0.844	0.131	0.156
Overall Test of Equality: 0.101			

Notes: The table shows the p-value from the test of the null hypothesis that $COV(Y^k, D_j)$ are equal in the sample and the simulation generated from our model. "Overall Test of Equality" reports the p-value for the the joint test against the null that all of the covariances for all outcomes and schooling levels are equal to zero.

A.13 Decomposing the Correlation Between ρ and S: Are Those Who Go to School the Ones Who Benefit from It?

The correlation between ρ_i and S_i is one possible measure of the nature of the sorting of people into schooling by their gain from it – a topic Becker investigated in depth in his Woytinsky lecture (1967; 1991). We have already established that the distributions of returns differ across schooling levels and the returns across schooling levels are far from perfectly correlated. It is thus of interest to push our analysis a bit further and investigate the correlation of annualized returns with attained schooling levels. We consider this question for direct returns and for total returns inclusive of continuation values. Table A38 shows the correlations between educational choices and the node-specific annualized (direct terminal) gains $\frac{(Y_j - Y_{j-1})}{(q_j - q_{j-1})}$ as well as the overall correlation.[24]

The correlations between the ρ_j and S are shown in column 1. Columns 2 through 4 show the correlations between the individual treatment effects ρ_j and choices at node D_j. For columns 2 through 4, each correlation is estimated conditional on the population that makes it to the specific decision ($Q_j = 1$).

[24] Precise definitions are given at the base of Table A38.

Table A38: Correlation Between Annualized Returns and Educational Choices

	$Corr(\rho, S)$	$Corr(\rho_1, (1-D_1))$ Grad. HS	$Corr(\rho_2, 1-D_2))$ Enroll in College	$Corr(\rho_3, (1-D_3))$ Grad. College
Wage	0.069	0.011	-0.041	0.053
	(0.011)	(0.061)	(0.002)	(0.018)
PV Wage	-0.080	-0.193	-0.068	0.084
	(0.037)	(0.034)	(0.023)	(0.021)
Smoking	-0.082	0.202	-0.110	-0.034
	(0.069)	(0.171)	(0.030)	(0.063)
Health Limits Work	-0.102	-0.225	0.227	-0.065
	(0.064)	(0.029)	(0.022)	(0.056)

Notes: Let q_j be the years of schooling associated with node j. The annualized terminal node j return is $\rho_j := \frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}$ and we define $\rho = \frac{Y_j - Y_0}{q_j - q_0}$. Total years of schooling is $S = \sum_{j=1}^{\bar{s}} q_j D_j$. Note $D_1 = 1$ if individuals stop their education as a high school graduate. $D_2 = 1$ and $D_3 = 1$ denote stopping at some college and college, respectively. Standard errors are estimated using 200 bootstrap samples and show the standard deviation of the estimate across the samples.

The overall correlation and the correlation by node differ substantially. The general pattern is that, for wages, people sort on terminal gains although the effect is only strong for graduating college, and for most outcomes it is perverse for enroll in college ("some college" in the text). The sorting is *negative* for PV wages, except for college graduation. For smoking, the overall effect is negative, but it is positive for high school graduation. For health limits work, the correlations differ but are negative except for the anomalous correlation for some college.

Table A39 decomposes the correlation between S and ρ in a fashion similar to what is reported in Table A38, except we work with dynamic treatment effects (T_j) inclusive of continuation values. This better represents the gains that agents use to make choices rather than the benefit associated with the comparison between terminal outcomes at j and $j-1$. The patterns are roughly similar across the two tables. The correlations are consistently negative for smoking across all transitions. The strongest negative correlation for health limits work is for high school graduation. The correlation with some college is anomalous.[25] Using either terminal level treatment effects or dynamic treatment effects, sorting is generally positive, broadly consistent with a meritocratic society.

Table A39: Correlation Between Returns and Educational Choice Including Continuation Values

	$Corr(\rho, S)$	$Corr(T_1, (1-D_1))$ Grad. HS	$Corr(T_2, 1-D_2))$ Enroll in College	$Corr(T_3, (1-D_3))$ Grad. College
Wage	0.069	-0.007	0.011	0.053
	(0.030)	(0.062)	(0.030)	(0.033)
PV Wage	-0.080	-0.102	0.002	0.084
	(0.030)	(0.052)	(0.027)	(0.032)
Smoking	-0.082	-0.030	-0.298	-0.034
	(0.089)	(0.130)	(0.084)	(0.101)
Health Limits Work	-0.102	-0.089	0.162	-0.065
	(0.079)	(0.132)	(0.111)	(0.110)

Notes: Let q_j be the years of schooling associated with node j. The annualized terminal node j return is $\rho_j := \frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}$ and we define $\rho = \sum_{j=1}^{\bar{s}} \rho_j (1-D_j)$. Total years of schooling is $S = \sum_{j-1}^{\bar{s}} q_j (1-D_j)$. $1-D_1 = 1$ stopping at high school, with $1-D_2$ and $1-D_3$ denoting stopping at some college and college, respectively.

A.14 Estimated Treatment Effects

This section documents a full set of treatment effects for each of our outcomes.

A.14.1 Treatment Effects Across Final Schooling Levels

We first present the traditional treatment effects across adjacent final levels corresponding to Figure 3 in the text. We then present treatment effects inclusive of continuation values for those corresponding to Figure 4.

Tables A40–A43 report traditional treatment effects by final schooling level: $\text{ATE}_{s,s'}$ (shown in Figure 3), treatment on the treated ($\text{TT}_{s,s'}$), and treatment on the untreated

[25] The category is a catch-all for those who attend college for remedial education, those who seek certificates, those getting associate's degrees, and dropouts from four-year college programs.

($\text{TUT}_{s,s'}$). We also display the raw difference ("observed") also shown in Figure 3, and ATEs derived from our model but computed for the entire population ($\text{ATE}^{\dagger}$).

These tables show the gains from switching from one final schooling level to another. All education levels are compared to dropouts as well as the level of education directly below it for both branches of Figure 1. $\text{ATE}_{s,s'}$ is the ATE computed from our model over the entire population. The other treatment parameters are defined for populations at the two final schooling levels. The difference between $\text{ATE}^{\dagger}$ and ATE is a measure of how different the characteristics are for those in the general population from those at the indicated pair of final schooling states. The differences between TT and ATE are measures of sorting gains. The differences between TUT and ATE are measures of sorting losses. Thus, in Table A40, the characteristics of people at the node deciding between the GED and dropping out are substantially less favorable than those in the general population, but there are little sorting gains or losses for this pair of alternatives. At the same time, there are substantial sorting gains (and losses) for those choosing between graduating college and not completing college. Moreover, the characteristics of people at this margin of choice are far more favorable.

A.14.2 Treatment Effects Across Nodes (Including Continuation Values)

We next present the traditional treatment effects including continuation values.

We show two tables for each outcome analyzed for populations conditional on $Q_j = 1$. The first is in the format similar to that of Tables A40–A43 and shows the population-wide average treatment effect, the average treatment effect for those who reach the node, treatment on the treated, treatment on the untreated (conditional on making it to the decision), and the average marginal treatment effect. These results are shown for all four branches of Figure 1. Each treatment effect is further broken into low-ability and high-ability samples where low-ability individuals are in the bottom half of both cognitive and non-cognitive ability, while high-ability individuals are in the top half of both cognitive and non-cognitive individuals.

The second table for each outcome shows the various treatment effects (population-wide average treatment effect, average treatment effect for those who reach the node, treatment on the treated, treatment on the untreated [conditional on making it to the decision], and the average marginal treatment effect) and decomposes them into their total effect and their direct effect (excluding option value). This is shown for each educational node.

Table A40: The Effects of Education on Log Wages, by *Final Schooling Level* Using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$ATE^{\dagger}_{s,s'}$	$ATE_{s,s'}$	$TT_{s,s'}$	$TUT_{s,s'}$	OLS
GED vs. HS Dropout	0.14	0.12	0.06	0.06	0.06	0.05
		(0.08)	(0.05)	(0.05)	(0.05)	(0.04)
HS Graduate vs. HS Dropout	0.24	0.13*	0.12**	0.12**	0.11**	0.08*
		(0.05)	(0.04)	(0.05)	(0.04)	(0.03)
Some College vs. HS Dropout	0.37	0.21**	0.21**	0.21**	0.22**	0.14**
		(0.06)	(0.05)	(0.07)	(0.06)	(0.04)
Four-Year College Degree vs. HS Dropout	0.64	0.26**	0.27**	0.34**	0.15*	0.28**
		(0.07)	(0.07)	(0.10)	(0.07)	(0.04)
Some College vs. HS Graduate	0.13	0.09**	0.10**	0.07**	0.11**	0.06*
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	0.26	0.04	0.11**	0.14**	0.08**	0.13**
		(0.04)	(0.04)	(0.04)	(0.04)	(0.03)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each row compares the outcomes from a particular schooling level j and the HS dropout status or from a particular schooling level j and schooling level $j-1$. The column "Observed" displays the observed differences in the data. The column "$ATE_{s,s'}$" displays the average treatment effect for everyone at one of the two listed final schooling levels. $ATE^{\dagger}_{s,s'}$ is evaluated over the whole population. The column "$TT_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout status (or $j-1$) but computed from those individuals selecting j as their final schooling decision. The column "$TUT_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout (or $j-1$) status but computed only for those individuals selecting "HS dropout" (or $j-1$) as their final schooling decision. For the first four entries, "OLS" shows results from the OLS regression $Y = \sum_{j \in \mathcal{J}} D_j * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$, where $\boldsymbol{A}$ contains ASVAB scores, GPA, and minor risky behavior as proxies for abilities and $\boldsymbol{X}$ is the standard set of controls used in the paper and an intercept, and we report the b_j. The last two rows of "OLS" show results from the OLS regression $Y = \sum_{j \in \mathcal{J}} Q_{j+1} * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$.

Table A41: The Effects of Education on Log PV of Wages, by ***Final Schooling Level*** Using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$\text{ATE}^{\dagger}_{s,s'}$	$\text{ATE}_{s,s'}$	$\text{TT}_{s,s'}$	$\text{TUT}_{s,s'}$	OLS
GED vs. HS Dropout	0.17	-0.20*	-0.11	-0.14*	-0.08	-0.01
		(0.10)	(0.06)	(0.06)	(0.08)	(0.05)
HS Graduate vs. HS Dropout	0.49	-0.04	0.07	-0.01	0.30**	0.19**
		(0.08)	(0.06)	(0.08)	(0.05)	(0.04)
Some College vs. HS Dropout	0.64	0.04	0.15*	-0.06	0.45**	0.24**
		(0.08)	(0.07)	(0.10)	(0.08)	(0.04)
Four-Year College Degree vs. HS Dropout	0.98	0.09	0.06	-0.09	0.33**	0.39**
		(0.09)	(0.10)	(0.13)	(0.10)	(0.05)
Some College vs. HS Graduate	0.15	0.08*	0.09**	0.06*	0.11**	0.06
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	0.34	0.06	0.17**	0.22**	0.11*	0.15**
		(0.06)	(0.04)	(0.05)	(0.05)	(0.04)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each row compares the outcomes from a particular schooling level j and the HS dropout status or from a particular schooling level j and schooling level $j-1$. The column "Observed" displays the observed differences in the data. The column "$\text{ATE}_{s,s'}$" displays the average treatment effect for everyone at one of the two listed final schooling levels. $\text{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population. The column "$\text{TT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout status (or $j-1$) but computed from those individuals selecting j as their final schooling decision. The column "$\text{TUT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout (or $j-1$) status but computed only for those individuals selecting "HS dropout" (or $j-1$) as their final schooling decision. For the first four entries, "OLS" shows results from the OLS regression $Y = \sum_{j \in \mathcal{J}} D_j * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$, where $\boldsymbol{A}$ contains ASVAB scores, GPA, and minor risky behavior as proxies for abilities and $\boldsymbol{X}$ is the standard set of controls used in the paper and an intercept, and we report the b_j. The last two rows of "OLS" show results from the OLS regression $Y = \sum_{j \in \mathcal{J}} Q_{j+1} * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$.

Table A42: The Effects of Education on Daily Smoking, by *Final Schooling Level* Using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$\text{ATE}^{\dagger}_{s,s'}$	$\text{ATE}_{s,s'}$	$\text{TT}_{s,s'}$	$\text{TUT}_{s,s'}$	OLS
GED vs. HS Dropout	-0.05	0.04	0.02	0.01	0.02	-0.03
		(0.09)	(0.05)	(0.06)	(0.05)	(0.05)
HS Graduate vs. HS Dropout	-0.28	-0.16*	-0.20**	-0.18*	-0.26**	-0.24**
		(0.08)	(0.06)	(0.08)	(0.05)	(0.04)
Some College vs. HS Dropout	-0.34	-0.22**	-0.23**	-0.20*	-0.27**	-0.28**
		(0.09)	(0.07)	(0.10)	(0.08)	(0.05)
Four-Year College Degree vs. HS Dropout	-0.53	-0.38**	-0.38**	-0.36**	-0.41**	-0.47**
		(0.09)	(0.09)	(0.14)	(0.09)	(0.05)
Some College vs. HS Graduate	-0.05	-0.05*	-0.06*	-0.07*	-0.05	-0.04
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
Four-Year College Degree vs. Some College	-0.19	-0.16**	-0.17**	-0.18**	-0.17**	-0,19**
		(0.04)	(0.04)	(0.05)	(0.04)	(0.04)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each row compares the outcomes from a particular schooling level j and the HS dropout status or from a particular schooling level j and schooling level $j-1$. The column "Observed" displays the observed differences in the data. The column "$\text{ATE}_{s,s'}$" displays the average treatment effect for everyone at one of the two listed final schooling levels. $\text{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population. The column "$\text{TT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout status (or $j-1$) but computed from those individuals selecting j as their final schooling decision. The column "$\text{TUT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout (or $j-1$) status but computed only for those individuals selecting "HS dropout" (or $j-1$) as their final schooling decision. For the first four entries, "OLS" shows results from the OLS regression $Y = \sum_{j \in \mathcal{J}} D_j * b_j + \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\alpha}$, where $\boldsymbol{A}$ contains ASVAB scores, GPA, and minor risky behavior as proxies for abilities and $\boldsymbol{X}$ is the standard set of controls used in the paper and an intercept, and we report the b_j. The last two rows of "OLS" show results from the OLS regression $Y = \sum_{j \in \mathcal{J}} Q_{j+1} * b_j + \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\alpha}$.

Table A43: The Effects of Education on Health Limits Work, by *Final Schooling Level* Using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$\text{ATE}^{\dagger}_{s,s'}$	$\text{ATE}_{s,s'}$	$\text{TT}_{s,s'}$	$\text{TUT}_{s,s'}$	OLS
GED vs. HS Dropout	-0.01	-0.03	0.06	0.05	0.06	0.04
		(0.08)	(0.05)	(0.06)	(0.06)	(0.04)
HS Graduate vs. HS Dropout	-0.16	-0.13*	-0.11*	-0.13*	-0.06	-0.08**
		(0.07)	(0.05)	(0.06)	(0.04)	(0.03)
Some College vs. HS Dropout	-0.21	-0.15**	-0.16**	-0.16*	-0.16*	-0.10**
		(0.07)	(0.06)	(0.09)	(0.07)	(0.04)
Four Year College Degree vs. HS Dropout	-0.30	-0.20**	-0.20**	-0.21**	-0.18*	-0.15**
		(0.08)	(0.09)	(0.13)	(0.09)	(0.04)
Some College vs. HS Graduate	-0.05	-0.02	-0.03	-0.00	-0.04	-0.02
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Four Year College Degree vs. Some College	-0.09	-0.05	-0.06*	-0.07*	-0.06*	-0.05
		(0.05)	(0.03)	(0.03)	(0.04)	(0.03)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each row compares the outcomes from a particular schooling level j and the HS dropout status or from a particular schooling level j and schooling level $j-1$. The column "Observed" displays the observed differences in the data. The column "$\text{ATE}_{s,s'}$" displays the average treatment effect for everyone at one of the two listed final schooling levels. $\text{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population. The column "$\text{TT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout status (or $j-1$) but computed from those individuals selecting j as their final schooling decision. The column "$\text{TUT}_{s,s'}$" displays the average treatment effects associated with a particular schooling level j relative to the HS dropout (or $j-1$) status but computed only for those individuals selecting "HS dropout" (or $j-1$) as their final schooling decision. For the first four entries, "OLS" shows results from the OLS regression $Y = \sum_{j \in \mathcal{J}} D_j * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$, where $\boldsymbol{A}$ contains ASVAB scores, GPA, and minor risky behavior as proxies for abilities and $\boldsymbol{X}$ is the standard set of controls used in the paper and an intercept, and we report the b_j. The last two rows of "OLS" show results from the OLS regression $Y = \sum_{j \in \mathcal{J}} Q_{j+1} * b_j + \boldsymbol{X'\beta} + \boldsymbol{A'\alpha}$.

Table A44: The Effects of Education on Log Wages, by *Decision Node*

	%	$ATE_j{}^\dagger$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		0.09*	0.09*	0.09	0.10**	0.09**
		(0.06)	(0.06)	(0.07)	(0.03)	(0.03)
Low Ability	0.31	0.10**	0.10**	0.09*	0.10**	
		(0.04)	(0.04)	(0.05)	(0.04)	
High Ability	0.31	0.09	0.09	0.10	0.07	
		(0.11)	(0.11)	(0.11)	(0.07)	
B. **Getting a GED** vs. **HS Dropout**						
All		0.12	0.06	0.06	0.06	0.06
		(0.08)	(0.05)	(0.05)	(0.05)	(0.05)
Low Ability	0.61	0.02	0.01	0.01	0.02	
		(0.06)	(0.05)	(0.06)	(0.06)	
High Ability	0.06	0.23	0.24*	0.23*	0.25*	
		(0.15)	(0.11)	(0.11)	(0.11)	
C. College Enrollment vs. **HS Graduate**						
All		0.13**	0.13**	0.14**	0.13**	0.10**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Low Ability	0.22	0.10*	0.10**	0.08**	0.11**	
		(0.05)	(0.04)	(0.04)	(0.04)	
High Ability	0.38	0.17**	0.17**	0.18**	0.15**	
		(0.03)	(0.03)	(0.04)	(0.03)	
D. **Four-Year College Degree** vs. **Some College**						
All		0.04	0.11**	0.14**	0.08**	0.11**
		(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Low Ability	0.14	-0.08	-0.05	-0.04	-0.05	
		(0.07)	(0.05)	(0.06)	(0.05)	
High Ability	0.51	0.18**	0.19**	0.19**	0.18**	
		(0.04)	(0.05)	(0.05)	(0.04)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision (inclusive of continuation value). Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $ATE_j^\dagger$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node ($Q_j = 1$). The TT_j column presents the average effect for those who chose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, $AMTE_j$ presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated treatment effects conditional upon endowment levels. The high (low) ability group is defined as those individuals with cognitive and socio-emotional endowment above (below) the overall median. For each decision node we display the fraction of individuals with low and high ability levels visiting each node.

Table A45: The Effects of Education on Log Wages, by *Decision Node* (Total and Direct Effects)

	$\text{ATE}_j{}^\dagger$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)	TUT_j	(Dir)	AMTE_j	(Dir)
A. Graduating from HS vs. Dropping from HS	0.094*	0.036	0.094*	0.036	0.093	0.021	0.100**	0.089**	0.093**	0.087**
	(0.056)	(0.056)	(0.056)	(0.056)	(0.072)	(0.068)	(0.029)	(0.031)	(0.028)	(0.032)
C. College Enrollment vs. **HS Graduate**	0.126**	0.086**	0.134**	0.085**	0.140**	0.062	0.128**	0.109**	0.101**	0.077**
	(0.027)	(0.027)	(0.025)	(0.029)	(0.031)	(0.040)	(0.026)	(0.030)	(0.023)	(0.028)
D. **Four-Year College Degree** vs. **Some College**	0.044		0.114**		0.141**		0.079**		0.110**	
	(0.044)		(0.037)		(0.042)		(0.036)		(0.034)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $\text{ATE}_j^\dagger$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, AMTE_j presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated direct treatment effects (Dir), which shows the effect of attaining only the next schooling level and does not include the continuation value.

Table A46: The Effects of Education on Log PV of Wages, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		0.17**	0.17**	0.14**	0.29**	0.28**
		(0.06)	(0.06)	(0.07)	(0.04)	(0.04)
Low Ability	0.31	0.27**	0.27**	0.22**	0.35**	
		(0.04)	(0.04)	(0.05)	(0.05)	
High Ability	0.31	0.09	0.09	0.09	0.12	
		(0.11)	(0.11)	(0.11)	(0.09)	
B. **Getting a GED** vs. **HS Dropout**						
All		-0.20*	-0.11	-0.14*	-0.08	-0.14*
		(0.10)	(0.06)	(0.06)	(0.08)	(0.06)
Low Ability	0.61	-0.19*	-0.13*	-0.17*	-0.10	
		(0.07)	(0.08)	(0.08)	(0.08)	
High Ability	0.06	-0.21	-0.04	-0.06	0.01	
		(0.19)	(0.16)	(0.15)	(0.17)	
C. College Enrollment vs. **HS Graduate**						
All		0.14**	0.14**	0.14**	0.14**	0.11**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low Ability	0.22	0.09*	0.06	0.01	0.08*	
		(0.06)	(0.05)	(0.05)	(0.05)	
High Ability	0.38	0.21**	0.21**	0.21**	0.21**	
		(0.04)	(0.04)	(0.04)	(0.04)	
D. **Four-Year College Degree** vs. **Some College**						
All		0.06	0.17**	0.22**	0.11*	0.15**
		(0.06)	(0.04)	(0.05)	(0.05)	(0.04)
Low Ability	0.14	-0.11	-0.01	0.04	-0.04	
		(0.10)	(0.07)	(0.07)	(0.08)	
High Ability	0.51	0.23**	0.26**	0.28**	0.21**	
		(0.05)	(0.05)	(0.05)	(0.05)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision (inclusive of continuation value). Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $ATE_j^{\dagger}$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who chose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, $AMTE_j$ presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated treatment effects conditional upon endowment levels. The high (low) ability group is defined as those individuals with cognitive and socio-emotional endowment above (below) the overall median. For each decision node we display the fraction of individuals with low- and high-ability levels visiting each node.

Table A47: The Effects of Education on Log PV of Wages, by *Decision Node* (Total and Direct Effects)

	$ATE^{\dagger}_j$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)	TUT_j	(Dir)	$AMTE_j$	(Dir)
A. Graduating from HS vs. Dropping from HS	0.173**	0.114*	0.173**	0.114*	0.138**	0.067	0.295**	0.279**	0.282**	0.269**
	(0.059)	(0.057)	(0.059)	(0.057)	(0.071)	(0.072)	(0.039)	(0.042)	(0.041)	(0.042)
C. College Enrollment vs. **HS Graduate**	0.138**	0.077*	0.137**	0.059*	0.139**	0.015	0.136**	0.106**	0.112**	0.072**
	(0.033)	(0.029)	(0.029)	(0.031)	(0.031)	(0.039)	(0.031)	(0.033)	(0.031)	(0.031)
D. **Four-Year College Degree** vs. **Some College**	0.056		0.171**		0.222**		0.106*		0.146**	
	(0.057)		(0.040)		(0.048)		(0.047)		(0.042)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $ATE^{\dagger}_j$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, $AMTE_j$ presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated direct treatment effects (Dir), which shows the effect of attaining only the next schooling level and does not include the continuation value.

Table A48: The Effects of Education on Daily Smoking, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Dropping from HS vs. Graduating from HS						
All		-0.26**	-0.26**	-0.27**	-0.26**	-0.24**
		(0.06)	(0.06)	(0.07)	(0.04)	(0.03)
Low Ability	0.31	-0.29**	-0.29**	-0.30**	-0.29**	
		(0.04)	(0.04)	(0.05)	(0.04)	
High Ability	0.31	-0.25**	-0.25**	-0.25**	-0.15*	
		(0.11)	(0.11)	(0.11)	(0.07)	
B. **Getting a GED** vs. **HS Dropout**						
All		0.04	0.02	0.01	0.02	0.01
		(0.09)	(0.05)	(0.06)	(0.05)	(0.05)
Low Ability	0.61	-0.00	0.00	-0.00	0.01	
		(0.05)	(0.05)	(0.05)	(0.06)	
High Ability	0.06	0.08	0.07	0.07	0.07	
		(0.17)	(0.14)	(0.14)	(0.13)	
C. College Enrollment vs. **HS Graduate**						
All		-0.12**	-0.14**	-0.18**	-0.10**	-0.13**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low Ability	0.22	-0.06	-0.09*	-0.12**	-0.07	
		(0.06)	(0.05)	(0.05)	(0.05)	
High Ability	0.38	-0.18**	-0.19**	-0.21**	-0.13**	
		(0.04)	(0.04)	(0.04)	(0.03)	
D. **Four-Year College Degree** vs. **Some College**						
All		-0.16**	-0.17**	-0.18**	-0.17**	-0.17**
		(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Low Ability	0.14	-0.12	-0.12*	-0.12	-0.11*	
		(0.08)	(0.07)	(0.08)	(0.07)	
High Ability	0.51	-0.20**	-0.19**	-0.19**	-0.20**	
		(0.05)	(0.05)	(0.05)	(0.04)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $ATE_j^{\dagger}$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, $AMTE_j$ presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated treatment effects conditional upon endowment levels. The high (low) ability group is defined as those individuals with cognitive and socio-emotional endowment above (below) the overall median. For each decision node we display the fraction of individuals with low- and high-ability levels visiting each node.

Table A49: The Effects of Education on Daily Smoking, by ***Decision Node*** (Total and Direct Effects)

	$\text{ATE}_j^{\dagger}$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)	TUT_j	(Dir)	AMTE_j	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.263**	-0.189**	-0.263**	-0.189**	-0.265**	-0.174**	-0.255**	-0.240**	-0.242**	-0.234**
	(0.056)	(0.058)	(0.056)	(0.058)	(0.071)	(0.069)	(0.036)	(0.038)	(0.033)	(0.038)
C. College Enrollment vs. **HS Graduate**	-0.115**	-0.054*	-0.139**	-0.065*	-0.178**	-0.080*	-0.096**	-0.049	-0.131**	-0.065**
	(0.031)	(0.031)	(0.028)	(0.033)	(0.033)	(0.045)	(0.032)	(0.037)	(0.027)	(0.033)
D. **Four-Year College Degree** vs. **Some College**	-0.164**		-0.172**		-0.176**		-0.167**		-0.173**	
	(0.044)		(0.043)		(0.051)		(0.038)		(0.038)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision (inclusive of continuation value). Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $\text{ATE}_j^{\dagger}$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, AMTE_j presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated direct treatment effects (Dir), which shows the effect of attaining only the next schooling level and does not include the continuation value.

Table A50: The Effects of Education on Health Limits Work, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		-0.11**	-0.11**	-0.11**	-0.09**	-0.11**
		(0.04)	(0.04)	(0.05)	(0.03)	(0.03)
Low Ability	0.31	-0.08**	-0.08**	-0.09**	-0.07	
		(0.04)	(0.04)	(0.04)	(0.04)	
High Ability	0.31	-0.11*	-0.11*	-0.11*	-0.12**	
		(0.07)	(0.07)	(0.07)	(0.06)	
B. **Getting a GED** vs. **HS Dropout**						
All		-0.03	0.06	0.05	0.06	0.06
		(0.08)	(0.05)	(0.06)	(0.06)	(0.05)
Low Ability	0.61	0.06	0.09*	0.09	0.10	
		(0.06)	(0.05)	(0.05)	(0.06)	
High Ability	0.06	-0.11	-0.05	-0.06	-0.04	
		(0.15)	(0.12)	(0.12)	(0.13)	
C. College Enrollment vs. **HS Graduate**						
All		-0.05	-0.04*	-0.02	-0.05*	-0.03
		(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Low Ability	0.22	-0.08	-0.07	-0.04	-0.08	
		(0.05)	(0.04)	(0.03)	(0.04)	
High Ability	0.38	-0.02	-0.02	-0.02	-0.02	
		(0.02)	(0.02)	(0.02)	(0.03)	
D. **Four-Year College Degree** vs. **Some College**						
All		-0.05	-0.06*	-0.07*	-0.06*	-0.07*
		(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Low Ability	0.14	-0.01	-0.02	-0.02	-0.01	
		(0.09)	(0.06)	(0.06)	(0.06)	
High Ability	0.51	-0.09**	-0.08*	-0.08*	-0.09**	
		(0.04)	(0.04)	(0.04)	(0.04)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $ATE_j^{\dagger}$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, $AMTE_j$ presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated treatment effects conditional upon endowment levels. The high (low) ability group is defined as those individuals with cognitive and socio-emotional endowment above (below) the overall median. For each decision node we display the fraction of individuals with low- and high-ability levels visiting each node.

Table A51: The Effects of Education on Health Limits Work, by ***Decision Node*** (Total and Direct Effects)

	$\text{ATE}_j^{\dagger}$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)	TUT_j	(Dir)	AMTE_j	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.108**	-0.097**	-0.108**	-0.097**	-0.113**	-0.101**	-0.090**	-0.084**	-0.110**	-0.104**
	(0.042)	(0.045)	(0.042)	(0.045)	(0.050)	(0.054)	(0.034)	(0.036)	(0.031)	(0.033)
C. College Enrollment vs. **HS Graduate**	-0.046	-0.023	-0.037*	-0.009	-0.023	0.016	-0.053*	-0.035	-0.029	-0.005
	(0.025)	(0.028)	(0.022)	(0.024)	(0.019)	(0.029)	(0.027)	(0.032)	(0.022)	(0.025)
D. **Four-Year College Degree** vs. **Some College**	-0.051		-0.064*		-0.070*		-0.057*		-0.067*	
	(0.048)		(0.031)		(0.034)		(0.036)		(0.030)	

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples. Each column presents the average effect of an educational decision. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. $\text{ATE}_j^{\dagger}$ represents the average effect for the full population, while ATE_j presents the average effect for those who visit the decision node. The TT_j column presents the average effect for those who choose a higher level of schooling ($D_j = 0$), and TUT_j presents the average effect for those who do not choose a higher level of schooling ($D_j = 1$). Finally, AMTE_j presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated direct treatment effects (Dir), which shows the effect of attaining only the next schooling level and does not include the continuation value.

A.15 Decomposing Observed Differences into Average Treatment Effects, Sorting Gains, and Selection Bias

We report decompositions conditional on final schooling level A.15.1 and by arrival at j, including continuation values A.15.2.

A.15.1 Decompositions by Final Schooling Level

We first decompose observed differences by final schooling level. We then decompose effects defined on $Q_j = 1$ that include continuation values.

Equation (20) can be written explicitly in terms of $\boldsymbol{X}$ and $\boldsymbol{\theta}$ as follows:

$$\begin{aligned}
&\underbrace{E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S \in \{j, j+1\}] + E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S \in \{j, j+1\}]}_{\text{ATE}} \\
&+ \underbrace{\Big(E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S = j+1] + E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S = j+1] - E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S \in \{j, j+1\}] - E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S \in \{j, j+1\}]\Big)}_{\text{Sorting Gains}} \\
&+ \underbrace{\Big(E[\tau^k_j(\boldsymbol{X})|S = j+1] + E[\boldsymbol{\theta}'\boldsymbol{\alpha}^k_j|S = j+1] - [E[\tau^k_j(\boldsymbol{X})|S = j] + E[\boldsymbol{\theta}'\boldsymbol{\alpha}^k_j|S = j]]\Big)}_{\text{Selection Bias}}. \qquad \text{(A.7)}
\end{aligned}$$

A.15.2 Decompositions: The Pairwise Observed Differences by Final Schooling Level

Table A52: Decomposition of the Observed Difference in Log Wages (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	0.25	0.12	0.00	0.13
	(0.03)	(0.04)	(0.01)	(0.05)
SC-HS	0.14	0.10	-0.03	0.07
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	0.23	0.11	0.03	0.09
	(0.03)	(0.04)	(0.02)	(0.03)
GED-DO	0.15	0.06	0.00	0.09
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A53: Decomposition of the Observed Difference in PV Log Wages (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	0.50	0.07	-0.08	0.51
	(0.04)	(0.06)	(0.02)	(0.08)
SC-HS	0.15	0.09	-0.03	0.09
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	0.31	0.17	0.05	0.09
	(0.04)	(0.04)	(0.02)	(0.04)
GED-DO	0.20	-0.11	-0.04	0.34
	(0.06)	(0.06)	(0.04)	(0.06)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A54: Decomposition of the Observed Difference in Smoking (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	-0.27	-0.20	0.02	-0.09
	(0.04)	(0.07)	(0.02)	(0.08)
SC-HS	-0.06	-0.06	-0.01	0.01
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	-0.19	-0.17	-0.00	-0.02
	(0.03)	(0.04)	(0.02)	(0.04)
GED-DO	-0.03	0.02	-0.01	-0.04
	(0.05)	(0.05)	(0.03)	(0.05)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A55: Decomposition of the Observed Difference in Health Limits Work (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	-0.17	-0.11	-0.02	-0.04
	(0.03)	(0.05)	(0.02)	(0.07)
SC-HS	-0.04	-0.03	0.02	-0.04
	(0.03)	(0.03)	(0.02)	(0.01)
Coll-SC	-0.09	-0.06	-0.01	-0.02
	(0.03)	(0.03)	(0.01)	(0.02)
GED-DO	-0.02	0.06	-0.01	-0.08
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Decomposing the Components into Observed Characteristics and Latent Ability (wage outcomes only)

Table A56: Decomposition of the Observed Difference in Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	0.25	0.12	0.13	-0.01	0.00	-0.01	0.01	0.13	0.10	0.02
	(0.03)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.03)	(0.04)
SC-HS	0.14	0.10	0.09	0.01	-0.03	-0.01	-0.01	0.07	0.04	0.03
	(0.04)	(0.05)	(0.07)	(0.06)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
Coll-SC	0.23	0.11	0.04	0.07	0.03	-0.00	0.03	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	0.15	0.06	0.14	-0.08	0.00	-0.01	0.01	0.09	0.05	0.04
	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A57: Decomposition of the Observed Difference in PV Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	0.50	0.07	-0.00	0.07	-0.08	-0.03	-0.04	0.51	0.29	0.22
	(0.04)	(0.06)	(0.08)	(0.02)	(0.02)	(0.01)	(0.02)	(0.08)	(0.05)	(0.06)
SC-HS	0.15	0.09	0.08	0.01	-0.03	-0.02	-0.01	0.09	0.06	0.03
	(0.06)	(0.06)	(0.09)	(0.07)	(0.04)	(0.02)	(0.03)	(0.06)	(0.05)	(0.04)
Coll-SC	0.31	0.17	0.09	0.09	0.05	0.01	0.04	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
GED-DO	0.20	-0.11	-0.12	0.01	-0.04	-0.03	-0.00	0.34	0.15	0.19
	(0.04)	(0.04)	(0.05)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A58: Decomposition of the Observed Difference in Smoking (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	-0.27	-0.20	-0.16	-0.03	0.02	-0.01	0.02	-0.09	0.03	-0.14
	(0.04)	(0.07)	(0.09)	(0.03)	(0.02)	(0.01)	(0.01)	(0.08)	(0.04)	(0.08)
SC-HS	-0.06	-0.06	-0.06	0.00	-0.01	-0.01	-0.00	0.01	0.02	-0.01
	(0.05)	(0.05)	(0.07)	(0.07)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	(0.05)
Coll-SC	-0.19	-0.17	-0.16	-0.03	-0.00	0.01	-0.01	-0.02	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	-0.03	0.02	0.04	-0.03	-0.01	-0.00	-0.00	-0.04	0.01	-0.06
	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A59: Decomposition of the Observed Difference in Health Limits Work (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	-0.17	-0.11	-0.12	0.02	-0.02	0.00	-0.02	-0.04	-0.04	-0.00
	(0.03)	(0.05)	(0.07)	(0.02)	(0.02)	(0.01)	(0.01)	(0.07)	(0.04)	(0.06)
SC-HS	-0.04	-0.03	-0.02	-0.00	0.02	0.01	0.02	-0.04	-0.01	-0.03
	(0.04)	(0.05)	(0.08)	(0.07)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)
Coll-SC	-0.09	-0.06	-0.05	-0.02	-0.01	0.00	-0.01	-0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
GED-DO	-0.02	0.06	-0.01	0.07	-0.01	-0.00	-0.00	-0.08	-0.03	-0.05
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

We further decompose the wage and PV effects into components of ability (cognitive and non-cognitive). See Tables A60 and A61.

Table A60: Decomposition of the Observed Difference in Wages (fully decomposed pairwise comparison)

		Average Treatment Effects				Sorting on Gains				Selection Bias			
	Observed	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog
HS-DO	0.25	0.12	0.13	-0.01	-0.00	0.00	-0.01	0.01	0.00	0.13	0.10	0.05	-0.03
	(0.03)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.03)	(0.03)	(0.02)
SC-HS	0.14	0.10	0.09	0.00	0.00	-0.03	-0.01	-0.02	0.01	0.07	0.04	0.04	-0.01
	(0.04)	(0.05)	(0.07)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.02)	(0.01)
Coll-SC	0.23	0.11	0.04	0.06	0.01	0.03	-0.00	0.03	0.00	0.09	0.07	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
GED-DO	0.15	0.06	0.14	-0.02	-0.06	0.00	-0.01	0.01	0.00	0.09	0.05	0.04	-0.00
	(0.03)	(0.04)	(0.04)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the “Observed” column for each row. The “Obs” and “Abil” columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table A61: Decomposition of the Observed Difference in PV Wages (fully decomposed pairwise comparison)

		Average Treatment Effects				Sorting on Gains				Selection Bias			
	Observed	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog
HS-DO	0.50	0.07	-0.00	0.06	0.01	-0.08	-0.03	-0.03	-0.01	0.51	0.29	0.22	0.01
	(0.04)	(0.06)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.08)	(0.05)	(0.05)	(0.03)
SC-HS	0.15	0.09	0.08	0.00	0.00	-0.03	-0.02	-0.02	0.01	0.09	0.06	0.04	-0.01
	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Coll-SC	0.31	0.17	0.09	0.05	0.04	0.05	0.01	0.02	0.02	0.09	0.07	0.03	-0.00
	(0.04)	(0.04)	(0.05)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)	(0.01)
GED-DO	0.20	-0.11	-0.12	0.01	0.00	-0.04	-0.03	-0.00	-0.00	0.34	0.15	0.19	0.00
	(0.06)	(0.06)	(0.09)	(0.04)	(0.06)	(0.04)	(0.02)	(0.03)	(0.01)	(0.06)	(0.05)	(0.04)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the “Observed” column for each row. The “Obs” and “Abil” columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

A.15.3 Decompositions in Observed Differences of Arriving at $j(Q_j = 1)$ Including Continuation Values

Parallel to the decomposition (20) in the text, we decompose the values of being at j into components associated with stopping at j and continuing beyond j where, for the upper branch of Figure 1 ($D_0 = 0$),

$$Y^k = Y_0^k + \sum_{j \geq 1}^{\bar{s}} \rho_{j-1,j}^k Q_j, \tag{A.8}$$

where $\rho_{j-1,j}^k = Y_j^k - Y_{j-1}^k$. The expected future gain for a person at j (≥ 1) is

$$\begin{aligned} & E_j \left(\sum_{l>j}^{\bar{s}} \rho_{l-1,l}^k Q_l | Q_j = 1 \right) \\ & = \sum_{l>j} \left[E_j(\rho_{l-1,l}^k | Q_l = 1) P(Q_l = 1 | Q_j = 1) \right], \quad j \geq 1, \end{aligned}$$

where the conditioning $D_0 = 0$ is kept implicit.

Introducing D_0 and noting that at the initial node, $Q_0 := 1$ and

$$Y^k = Y_0 + \left(\sum_{j \in \mathcal{S} \setminus \{0,G\}} \rho_{j-1,j}^k Q_j \right) (1 - D_0) + \rho_{0,G}^k Q_G(D_0),$$

where $\rho_{0,G}^k = (Y_G^k - Y_0^k)$. Thus, the expected future gain for a person at $j = 0$ is

$$\begin{aligned} & E_0 \left[\left(\sum_{l \geq 1} \rho_{l-1,l}^k Q_l (1 - D_0) | Q_l = 1, D_0 = 0 \right) P(Q_l = 1 | D_0 = 0) + \rho_{0,G}^k Q_G (1 - D_0) \right] \qquad (A.9) \\ & = \sum_{l \geq 1} E_0 \left(\rho_{l-1,l}^k | Q_l = 1, D_0 = 0 \right) P(Q_l = 1 | D_0 = 0) + E_0 \left(\rho_{0,G}^k | Q_G = 1, D_0 = 1 \right) P(Q_G = 1 | D_0 = 1). \end{aligned}$$

Specifically for the kth outcome at node j:

$$
\begin{aligned}
&\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Observed difference}}\\
=&\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=0,Q_j=1,Fix\ D_j=1]}_{\text{Dynamic treatment on the treated for those at } j}\\
+&\underbrace{E[Y^k|D_j=0,Q_j=1,Fix\ D_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Selection bias for those at } j}\\
=&\underbrace{E[Y^k|Q_j=1,Fix\ D_j=0]-E[Y^k|Q_j=1,Fix\ D_j=1]}_{\text{ATE for those at } j}\\
+&\underbrace{\left\{\begin{matrix}\left(E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=0,Q_j=1,Fix\ D_j=1]\right)\\ -\left(E[Y^k|Q_j=1,Fix\ D_j=0]-E[Y^k|Q_j=1,Fix\ D_j=1]\right)\end{matrix}\right\}}_{\text{TT - ATE: Sorting gain at } j \text{ for those who transit to } j+1}\\
+&\underbrace{E[Y^k|D_j=0,Q_j=1,Fix\ D_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Selection bias}}.
\end{aligned}
\tag{A.10}
$$

The node-specific ATE_j is defined for the population at $Q_j=1$ and considers moving the entire group from j to $j+1$ (i.e, $Fix\ D_j=1$ and $Fix\ D_j=0$, respectively). The sorting gain is the net gain beyond ATE_j to those who actually take the transition ($D_j=0$).

Table A62: Decomposition of the Observed Difference in Wage (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	0.32	0.09	-0.00	0.22
	(0.02)	(0.06)	(0.02)	(0.07)
Enroll in Coll.	0.27	0.13	0.01	0.13
	(0.02)	(0.03)	(0.01)	(0.02)
Graduate Coll.	0.23	0.11	0.03	0.09
	(0.03)	(0.04)	(0.02)	(0.03)
Get GED	0.15	0.06	0.00	0.09
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table A63: Decomposition of the Observed Difference in PV Wage (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	0.58	0.17	-0.04	0.44
	(0.03)	(0.06)	(0.02)	(0.07)
Enroll in Coll.	0.32	0.14	0.00	0.18
	(0.03)	(0.03)	(0.02)	(0.03)
Graduate Coll.	0.31	0.17	0.05	0.09
	(0.04)	(0.04)	(0.02)	(0.04)
Get GED	0.20	-0.11	-0.04	0.34
	(0.06)	(0.06)	(0.04)	(0.06)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table A64: Decomposition of the Observed Difference in Smoking (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	-0.34	-0.26	-0.00	-0.08
	(0.03)	(0.06)	(0.01)	(0.06)
Enroll in Coll.	-0.16	-0.14	-0.04	0.01
	(0.02)	(0.03)	(0.02)	(0.02)
Graduate Coll.	-0.19	-0.17	-0.00	-0.02
	(0.03)	(0.04)	(0.02)	(0.04)
Get GED	-0.03	0.02	-0.01	-0.04
	(0.05)	(0.05)	(0.03)	(0.05)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table A65: Decomposition of the Observed Difference in Health Limits Work (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	-0.21	-0.11	-0.00	-0.09
	(0.02)	(0.04)	(0.01)	(0.04)
Enroll in Coll.	-0.09	-0.04	0.01	-0.07
	(0.02)	(0.02)	(0.01)	(0.02)
Graduate Coll.	-0.09	-0.06	-0.01	-0.02
	(0.02)	(0.03)	(0.01)	(0.02)
Get GED	-0.02	0.06	-0.01	-0.08
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains, and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

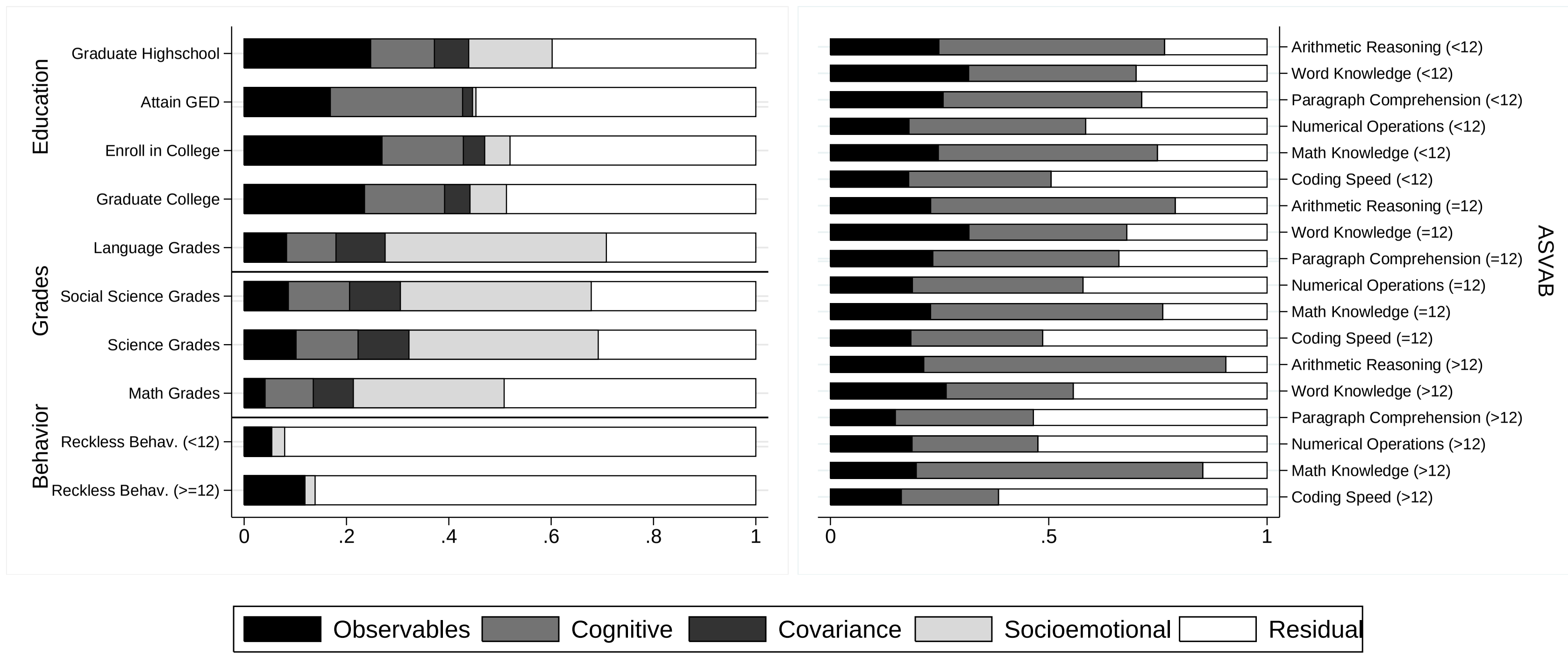
A.16 Variance Decompositions

The variance of the endowments is split up into its components: $var(\alpha_C\theta_C + \alpha_{SE}\theta_{SE}) = \alpha_C^2\sigma_{\theta_C}^2 + 2 * \alpha_C\alpha_{SE}cov(\theta_C, \theta_{SE}) + \alpha_{SE}^2\sigma_{\theta_{SE}}^2$. These are represented in the table as Cognitive ($\alpha_C^2\sigma_{\theta_C}^2$), Covariance ($2 * \alpha_C\alpha_{SE}cov(\theta_C, \theta_{SE})$), and Socio-emotional ($\alpha_{SE}^2\sigma_{\theta_{SE}}^2$). For discrete variables, the decomposition is for the index generating outcome.

Figure A15 below decomposes the variance in our measurement system into (i) variance explained by observables; (ii) variance explained by our unobserved factors; and (iii) the remaining unexplained variance. We further decompose the variance explained by the unobserved factors into the unique cognitive component, the unique non-cognitive component, and the component that cannot be assigned due to positive correlation between the factors.

The observables explain around 20 to 30% of the variance in educational choice while the factors account for 30 to 40% of the variance. For high school graduation, cognitive and socio-emotional endowments are of about equal importance. For college enrollment, both factors matter, but cognition plays a bigger role. For GED certification, only cognition matters with very little of the variance being explained by socio-emotional endowment. Observable characteristics explain 10 to 15% of grades while the factors explain up to 60% of the variance in grades. Socio-emotional endowments explain substantially more of the variance than cognition. Observables and factors explain from 40 to almost 90% of the variance in test scores. Observable characteristics explain 20 to 30% while the cognitive factor explains 20 to 65%.

Figure A15: 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 socio-emotional 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 characteristics, 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 "Socio-Emotional" is $var(\theta_{NC}\alpha_{NC})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$. The ASVAB tests are assumed to not depend on the socio-emotional 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.

Table A66: Variance Decomposition of Educational Decisions and Grades

	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
D_1: Graduate HS	0.247	0.122	0.068	0.164	0.399
D_2: Enroll College	0.270	0.159	0.043	0.050	0.478
D_3: Graduate College	0.238	0.155	0.051	0.073	0.483
GPA Language	0.084	0.094	0.097	0.434	0.292
GPA Social Sciences	0.086	0.116	0.100	0.375	0.322
GPA Science	0.102	0.118	0.101	0.372	0.308
GPA Math	0.041	0.091	0.079	0.296	0.493

Notes: Columns show the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socio-emotional 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 characteristics, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\mathbf{X'}\boldsymbol{\beta}) + var(\boldsymbol{\theta'}\boldsymbol{\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\mathbf{X'}\boldsymbol{\beta}) + var(\boldsymbol{\theta'}\boldsymbol{\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha'}\boldsymbol{\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE}) + var(\theta_{SE}\alpha_{SE})$. In the legend above, for continuous outcomes, “Observables” is $var(\mathbf{X'}\boldsymbol{\beta})/var(Y)$, “Cognitive” is $var(\theta_C\alpha_C)/var(Y)$, “Covariance” is $2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE})/var(Y)$, and “Socio-Emotional” is $var(\theta_{SE}\alpha_{SE})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$.

Table A67: Variance Decomposition of ASVAB Tests

	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
Arithmetic Reasoning (< 12)	0.248	0.517	0.000	0.000	0.235
Word Knowledge (< 12)	0.316	0.383	0.000	0.000	0.301
Paragraph Comprehension (< 12)	0.257	0.455	0.000	0.000	0.288
Numerical Operations (< 12)	0.180	0.404	0.000	0.000	0.416
Math Knowledge (< 12)	0.247	0.501	0.000	0.000	0.252
Coding Speed (< 12)	0.179	0.326	0.000	0.000	0.495
Arithmetic Reasoning (= 12)	0.229	0.559	0.000	0.000	0.212
Word Knowledge (= 12)	0.316	0.361	0.000	0.000	0.323
Paragraph Comprehension (= 12)	0.235	0.426	0.000	0.000	0.339
Numerical Operations (= 12)	0.188	0.389	0.000	0.000	0.423
Math Knowledge (= 12)	0.230	0.530	0.000	0.000	0.240
Coding Speed (= 12)	0.183	0.301	0.000	0.000	0.515
Arithmetic Reasoning (> 12)	0.212	0.695	0.000	0.000	0.093
Word Knowledge (> 12)	0.262	0.291	0.000	0.000	0.447
Paragraph Comprehension (> 12)	0.147	0.316	0.000	0.000	0.536
Numerical Operations (> 12)	0.185	0.288	0.000	0.000	0.527
Math Knowledge (> 12)	0.196	0.656	0.000	0.000	0.148
Coding Speed (> 12)	0.161	0.222	0.000	0.000	0.616

Notes: Columns show the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socio-emotional 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 characteristics, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha}'\boldsymbol{\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE}) + var(\theta_{SE}\alpha_{SE})$. In the legend above, for continuous outcomes, "Observables" is $var(\mathbf{X}'\boldsymbol{\beta})/var(Y)$, "Cognitive" is $var(\theta_C\alpha_C)/var(Y)$, "Covariance" is $2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE})/var(Y)$, and "Socio-Emotional" is $var(\theta_{SE}\alpha_{SE})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$.

Table A68: Variance Decomposition of Early Reckless and Adverse Behaviors

	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
Early Reckless (9th-11th)[b]	0.053	0.000	-0.001	0.027	0.921
Early Reckless (12th)[b]	0.116	0.001	-0.002	0.022	0.863
Early Marijuana[c]	0.077	0.004	0.010	0.126	0.783
Early Daily Smoking[c]	0.068	0.016	0.019	0.095	0.803
Early Drinking[c]	0.022	0.008	0.008	0.033	0.930
Early Intercourse[c]	0.114	0.028	0.019	0.054	0.785

Notes: Columns show the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socio-emotional 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 characteristics, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha}'\boldsymbol{\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE}) + var(\theta_{SE}\alpha_{SE})$. In the legend above, for continuous outcomes, "Observables" is $var(\mathbf{X}'\boldsymbol{\beta})/var(Y)$, "Cognitive" is $var(\theta_C\alpha_C)/var(Y)$, "Covariance" is $2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE})/var(Y)$, and "Socio-Emotional" is $var(\theta_{SE}\alpha_{SE})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$.

Table A69: Variance Decomposition of Outcomes

	Observables	Cognitive	Covariance	Socio-Emotional	Unobservables
Log Wages (30)	0.188	0.082	0.010	0.006	0.715
High school dropouts	0.284	0.027	-0.007	0.009	0.688
High school graduates	0.180	0.059	-0.008	0.005	0.764
Some college	0.176	0.004	0.001	0.002	0.817
Four-year college graduate	0.105	0.125	0.014	0.007	0.749
PV Log Wages (30)	0.251	0.091	0.013	0.008	0.637
High school dropouts	0.442	0.162	0.003	0.000	0.393
High school graduates	0.272	0.050	-0.012	0.011	0.678
Some college	0.189	0.011	-0.001	0.000	0.800
Four-year college graduate	0.135	0.083	0.023	0.027	0.731
Smoking Age 30	0.050	0.042	0.027	0.076	0.804
High school dropouts	0.167	0.038	0.023	0.059	0.713
High school graduates	0.068	0.000	0.001	0.007	0.924
Some college	0.099	0.002	0.002	0.009	0.888
Four-year college graduate	0.097	0.021	0.015	0.046	0.822
Health Limits Work	0.060	0.046	0.014	0.020	0.860
High school dropouts	0.116	0.037	-0.018	0.039	0.826
High school graduates	0.076	0.038	0.008	0.008	0.870
Some college	0.071	0.000	0.000	0.008	0.921
Four-year college graduate	0.125	0.045	0.007	0.005	0.817

Notes: Columns show the fraction of the variance in each outcome explained by observable covariates (X), unobservable cognitive and socio-emotional 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 characteristics, and unobservables are all independent, the total variance of an outcome can be decomposed as $var(Y) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for continuous outcomes and $var(I) = var(\mathbf{X}'\boldsymbol{\beta}) + var(\boldsymbol{\theta}'\boldsymbol{\alpha}) + var(\epsilon)$ for discrete outcomes. Furthermore, $var(\boldsymbol{\alpha}'\boldsymbol{\theta}) = var(\theta_C\alpha_C) + 2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE}) + var(\theta_{SE}\alpha_{SE})$. In the legend above, for continuous outcomes, "Observables" is $var(\mathbf{X}'\boldsymbol{\beta})/var(Y)$, "Cognitive" is $var(\theta_C\alpha_C)/var(Y)$, "Covariance" is $2cov(\theta_C\alpha_C, \theta_{SE}\alpha_{SE})/var(Y)$, and "Socio-Emotional" is $var(\theta_{SE}\alpha_{SE})/var(Y)$. Calculations for the discrete outcomes are the same, but are normalized by $var(I)$ rather than $var(Y)$.

A.17 OLS Estimates of Treatment Effects and Continuation Values

A.17.1 Models with Continuation Values

We first analyze models with continuation values. We then consider models comparing pairs' final states, ignoring continuation values. Following Cameron and Heckman (1993), we estimate models for outcome Y^k for person i of the form:

$$Y^k = \sum_{j \in S} b_j^k Q_j + \boldsymbol{\theta}_i' \boldsymbol{\alpha}^k + \boldsymbol{X}_i' \boldsymbol{\beta}^k + \epsilon_i, \tag{A.11}$$

where b_j^k is the estimated gain from going to j from $j-1$. These correspond to the $E(\rho_{j-1,j}^k | Q_j = 1)$ defined in Section A.15.3.

We decompose these effects into the direct effects:

$$\widetilde{DE}_j^k = b_{j+1}^k$$

and the continuation value, which for the upper branch ($D_0 = 0$), is:

$$\tilde{C'}_{j+1}^k = \sum_{l > j+1}^{\bar{s}} b_l^k E(Q_l | D_j = 0).$$

The total effect is $\widetilde{DE}_j^k + \tilde{C}_{j+1}^k$. Table A70 presents the estimates from this approach to estimating both direct effects and continuation values where we condition on the $\boldsymbol{\theta}$ and $\boldsymbol{X}$ used in the main model. To compute this decomposition, we use a probit model to estimate the probability of each transition controlling for $\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}$ to generate $E(Q_l | D_j = 0)$.

Table A70 compares estimates from the main model of this paper to estimates from a variety of linear model OLS specifications. Three OLS models are considered. The first uses only age and age squared as regressors. The second uses the full set of control variables utilized to estimate the main model but excludes any proxies for cognitive and non-cognitive endowments. The third OLS model includes the full set of control variables as well as proxies for $\boldsymbol{\theta}$, including age-adjusted AFQT scores, 9th grade GPA, and an indicator of minor risky or reckless behavior. For each outcome, six numbers are reported. The first three rows report the dynamic treatment effects (Equation (13)) for each educational transition (inclusive of continuation values). The fifth and sixth numbers show the continuation values associated with graduating from high school and the continuation value of graduating from "some college," respectively. Transition probabilities are estimated using the same covariates as are used to predict outcomes, plus the full set of exclusion restriction instruments used in estimating the main model.

If proxies for cognitive and non-cognitive endowments are not included, or the full set of controls are not included, the OLS estimates are substantially different than the estimates from the main model. Typically, we find that the estimated treatment effects are much larger than those obtained from the main model. The OLS estimates with controls and proxies for abilities produces estimates much more similar to those estimated from the main model of this paper. Yet, the OLS results are not identical and are sometimes substantially different for certain treatment effects, such as the effect of graduating from high school on wages or PV wages.

Table A70: Estimated Dynamic Treatment Effects and Continuation Values Using OLS

		Very Simple		OLS-Simple		OLS		Main Model	
Wages	TE: HS-DO	0.321	(0.030)	0.269	(0.028)	0.172	(0.031)	0.094	(0.056)
	TE: SC-HS	0.183	(0.025)	0.154	(0.024)	0.098	(0.025)	0.134	(0.027)
	TE: CG-SC	0.274	(0.037)	0.218	(0.036)	0.164	(0.037)	0.114	(0.044)
	CV: HS-DO	0.080	(0.011)	0.077	(0.010)	0.056	(0.010)	0.058	(0.013)
	CV: SC-HS	0.069	(0.010)	0.056	(0.009)	0.042	(0.010)	0.050	(0.013)
PV Wages	TE: HS-DO	0.522	(0.047)	0.407	(0.044)	0.269	(0.047)	0.173	(0.059)
	TE: SC-HS	0.219	(0.030)	0.158	(0.027)	0.084	(0.029)	0.137	(0.033)
	TE: CG-SC	0.347	(0.039)	0.225	(0.038)	0.148	(0.040)	0.171	(0.057)
	CV: HS-DO	0.096	(0.013)	0.079	(0.011)	0.049	(0.012)	0.059	(0.014)
	CV: SC-HS	0.087	(0.010)	0.057	(0.010)	0.038	(0.010)	0.078	(0.014)
Health Limits Work	TE: HS-DO	-0.156	(0.036)	-0.142	(0.035)	-0.086	(0.040)	-0.108	(0.042)
	TE: SC-HS	-0.062	(0.024)	-0.057	(0.024)	-0.028	(0.026)	-0.037	(0.025)
	TE: CG-SC	-0.097	(0.026)	-0.078	(0.028)	-0.048	(0.030)	-0.064	(0.048)
	CV: HS-DO	-0.027	(0.010)	-0.028	(0.010)	-0.016	(0.011)	-0.010	(0.008)
	CV: SC-HS	-0.024	(0.007)	-0.020	(0.007)	-0.012	(0.008)	-0.028	(0.011)
Smoking	TE: HS-DO	-0.338	(0.041)	-0.363	(0.041)	-0.300	(0.045)	-0.263	(0.056)
	TE: SC-HS	-0.100	(0.027)	-0.123	(0.028)	-0.100	(0.029)	-0.139	(0.031)
	TE: CG-SC	-0.197	(0.037)	-0.231	(0.038)	-0.195	(0.039)	-0.172	(0.044)
	CV: HS-DO	-0.044	(0.012)	-0.063	(0.012)	-0.060	(0.012)	-0.075	(0.013)
	CV: SC-HS	-0.050	(0.009)	-0.059	(0.010)	-0.050	(0.010)	-0.074	(0.017)

Notes: "Very Simple" is estimated using ordinary least squares using only age and age squared as covariates."OLS-Simple" is estimated using OLS with the full set of controls, but without proxies for cognitive or non-cognitive ability. "OLS" is estimated using OLS adding age-adjusted AFQT, 9th grade GPA, and an indicator of early risky behavior as proxies for 0. "Main Model" are estimates from the model presented in the paper. The table shows the estimated dynamic treatment effects for each of the educational transitions ("TE") as well as the associated continuation value ("CV") from graduating from high school and enrolling in college. Results for GED certification are included in the model but not displayed. All standard errors are computed using 200 bootstrap samples.

A.17.2 Comparing Pairwise Treatment Effects by Final Schooling Level from OLS and Our Model

This section compares OLS estimates of the pairwise treatment effects by final schooling level ($\mathrm{ATE}_{s,s'}$) with estimates from our model.[26] We also compare a version of $\mathrm{ATE}_{s,s'}$ from our main model (labeled $\mathrm{ATE}^{\dagger}_{s,s'}$) that is computed over the entire population. The contrast between $\mathrm{ATE}_{s,s'}$ and $\mathrm{ATE}^{\dagger}_{s,s'}$ indicates the strength of compositional effects that arise from

[26] $\mathrm{ATE}_{s,s'}$ is defined in Equation (18).

comparing the populations at s, s' with the general population.

The OLS model we estimate is a linear projection approximation to

$$E(Y|Fix\ D_s = 0, D_s + D_{s'} = 1, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\theta})$$
$$-E(Y|Fix\ D_s = 0, D_s + D_{s'} = 1, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\theta}),$$

which, under conditional independence assumptions (A-1a)-(A-1e), is the same as

$$E(Y|D_s = 0, D_s + D_{s'} = 1, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$$
$$-E(Y|D_s = 1, D_s + D_{s'} = 1, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}).$$

This is the outcome associated with fixing $D_s = 0$ for a population at final level s or s' and subtracting the mean outcome associated with fixing $D_s = 1$ for the same population. Tables A71–A74 compare OLS estimates (including $\boldsymbol{\theta}$) with model estimates of $\text{ATE}_{s-1,s}$ (Equation (18)).

Table A71: Comparing Observed and OLS Estimates of $\mathrm{ATE}_{s,s'}$ and $\mathrm{ATE}^{\dagger}_{s,s'}$ to Our Model: Wages

	Observed	OLS	$\mathrm{ATE}^{\dagger}_{s,s'}$	$\mathrm{ATE}_{s,s'}$
GED vs. HS Dropout	0.13	0.05	0.12	0.06
		(0.04)	(0.08)	(0.05)
HS Graduate vs. HS Dropout	0.24	0.08*	0.13*	0.12**
		(0.03)	(0.05)	(0.04)
Some College vs. HS Graduate	0.13	0.06*	0.10**	0.07**
		(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	0.27	0.14**	0.04	0.11**
		(0.03)	(0.04)	(0.04)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples for the model. Each row compares the outcomes from a particular schooling level j and $j-1$. The column "OLS" displays the coefficient from an OLS regression controlling for the standard controls used in the paper and proxies for ability. The regression is of the form $Y = \sum_{j \in \mathcal{J}} Q_{j+1} b_j + \boldsymbol{X\beta} + \boldsymbol{A'\alpha}$ where $\boldsymbol{X}$ are the controls and $\boldsymbol{A}$ are the proxies of ability. The column "ATE" displays the average treatment effect obtained from the comparison of the outcomes associated with a particular schooling level j relative to $j-1$. $\mathrm{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population, whereas $\mathrm{ATE}_{s,s'}$ is evaluated for everyone with whose final schooling level is j or $j-1$.

Table A72: Comparing Observed and OLS Estimates of $\mathrm{ATE}_{s,s'}$ and $\mathrm{ATE}^{\dagger}_{s,s'}$ to Our Model: PV Wages

	Observed	OLS	$\mathrm{ATE}^{\dagger}_{s,s'}$	$\mathrm{ATE}_{s,s'}$
GED vs. HS Dropout	0.17	-0.01	-0.20*	-0.11
		(0.05)	(0.10)	(0.06)
HS Graduate vs. HS Dropout	0.49	0.19**	-0.04	0.07
		(0.04)	(0.08)	(0.06)
Some College vs. HS Graduate	0.15	0.06	0.08*	0.09**
		(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	0.34	0.15**	0.06	0.17**
		(0.04)	(0.06)	(0.04)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples for the model. Each row compares the outcomes from a particular schooling level j and $j-1$. The column "OLS" displays the coefficient from an OLS regression controlling for the standard controls used in the paper and proxies for ability. The regression is of the form $Y = \sum_{j \in \mathcal{J}} Q_{j+1} b_j + \boldsymbol{X\beta} + \boldsymbol{A'\alpha}$ where $\boldsymbol{X}$ are the controls and $\boldsymbol{A}$ are the proxies of ability. The column "ATE" displays the average treatment effect obtained from the comparison of the outcomes associated with a particular schooling level j relative to $j-1$. $\mathrm{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population, whereas $\mathrm{ATE}_{s,s'}$ is evaluated for everyone with whose final schooling level is j or $j-1$.

Table A73: Comparing Observed and OLS Estimates of $\mathbf{ATE}_{s,s'}$ and $\mathbf{ATE}^{\dagger}_{s,s'}$ to Our Model: Smoking

	Observed	OLS	$\text{ATE}^{\dagger}_{s,s'}$	$\text{ATE}_{s,s'}$
GED vs. HS Dropout	-0.05	-0.03	0.04	0.02
		(0.05)	(0.09)	(0.05)
HS Graduate vs. HS Dropout	-0.28	-0.24**	-0.16*	-0.20**
		(0.04)	(0.08)	(0.06)
Some College vs. HS Graduate	-0.05	-0.04	-0.05*	-0.06*
		(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	-0.19	-0.19**	-0.16**	-0.17**
		(0.04)	(0.04)	(0.04)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples for the model. Each row compares the outcomes from a particular schooling level j and $j-1$. The column "OLS" displays the coefficient from an OLS regression controlling for the standard controls used in the paper and proxies for ability. The regression is of the form $Y = \sum_{j \in \mathcal{J}} Q_{j+1} b_j + \boldsymbol{X\beta} + \boldsymbol{A'\alpha}$ where $\boldsymbol{X}$ are the controls and $\boldsymbol{A}$ are the proxies of ability. The column "ATE" displays the average treatment effect obtained from the comparison of the outcomes associated with a particular schooling level j relative to $j-1$. $\text{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population, whereas $\text{ATE}_{s,s'}$ is evaluated for everyone with whose final schooling level is j or $j-1$.

Table A74: Comparing Observed and OLS Estimates of $\mathbf{ATE}_{s,s'}$ and $\mathbf{ATE}^{\dagger}_{s,s'}$ to Our Model: Health Limits Work

	Observed	OLS	$\text{ATE}^{\dagger}_{s,s'}$	$\text{ATE}_{s,s'}$
GED vs. HS Dropout	-0.01	0.04	-0.03	0.06
		(0.04)	(0.08)	(0.05)
HS Graduate vs. HS Dropout	-0.16	-0.08**	-0.13*	-0.11*
		(0.03)	(0.07)	(0.05)
Some College vs. HS Graduate	-0.05	-0.03	-0.02	-0.03
		(0.03)	(0.03)	(0.03)
Four-Year College Degree vs. Some College	-0.09	-0.05	-0.05	-0.06*
		(0.03)	(0.05)	(0.03)

Notes: Standard errors (in parenthesis) and significance levels ($* = p < 0.05, ** = p < 0.01$) are calculated using 200 bootstrap samples for the model. Each row compares the outcomes from a particular schooling level j and $j-1$. The column "OLS" displays the coefficient from an OLS regression controlling for the standard controls used in the paper and proxies for ability. The regression is of the form $Y = \sum_{jinJ} Q_{j+1} b_j + \boldsymbol{X\beta} + \boldsymbol{A'\alpha}$ where $\boldsymbol{X}$ are the controls and $\boldsymbol{A}$ are the proxies of ability. The column "ATE" displays the average treatment effect obtained from the comparison of the outcomes associated with a particular schooling level j relative to $j-1$. $\text{ATE}^{\dagger}_{s,s'}$ is evaluated over the whole population, whereas $\text{ATE}_{s,s'}$ is evaluated for everyone with whose final schooling level is j or $j-1$.

The OLS and ATE estimates are in fairly close agreement for wages, smoking, and health limits work. However, the OLS and ATE estimates are generally further apart for log present value of wages. Comparing ATE with ATE†, we find large compositional effects for wage and log PV wage. Compositional effects are small for smoking, which is consistent with our evidence that there is little selection bias for that outcome. For health limits work, the compositional effects are larger than those for smoking, but smaller than those for wage outcomes. See Tables A52–A55 for the estimates of selection bias.

A.18 The Estimators Used to Generate the Estimates Reported in Table 3 in the Text

This subsection briefly discusses the estimators used to generate the alternative estimators reported in Table 3 in the paper. They are used to estimate the returns to making a particular educational transition (inclusive of continuation values). We consider both matching and OLS estimators that are designed to estimate

$$E(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, Q_j = 1, Fix\ D_j = 0) - E(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, Q_j = 1, Fix\ D_j = 1). \tag{A.12}$$

Because of assumptions (A-1a)-(A-1e), this expression is equivalent to

$$E(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, Q_j = 1, D_j = 0) - E(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, Q_j = 1, D_j = 1). \tag{A.13}$$

The linear regression estimators are straightforward OLS applied using the various sets of regressors and interactions explained in the note to Table 3. The column labeled "OLS" includes the control variables listed in Table 3 but excludes any proxy for $\boldsymbol{\theta}$. All linear regression estimators do not include the exclusion restrictions listed in Table 1.

The other three columns listed under the heading "Linear Regression" are from models that include the variables used to produce the estimates in the first column of Table 3, plus various measurements or proxies for $\boldsymbol{\theta}$. The second column is based on proxies for $\boldsymbol{\theta}$. The third and fourth columns are based on Bartlett (1937, 1938) score estimates of $\boldsymbol{\theta}$. The fourth column interacts treatment status with all control variables and Bartlett scores.

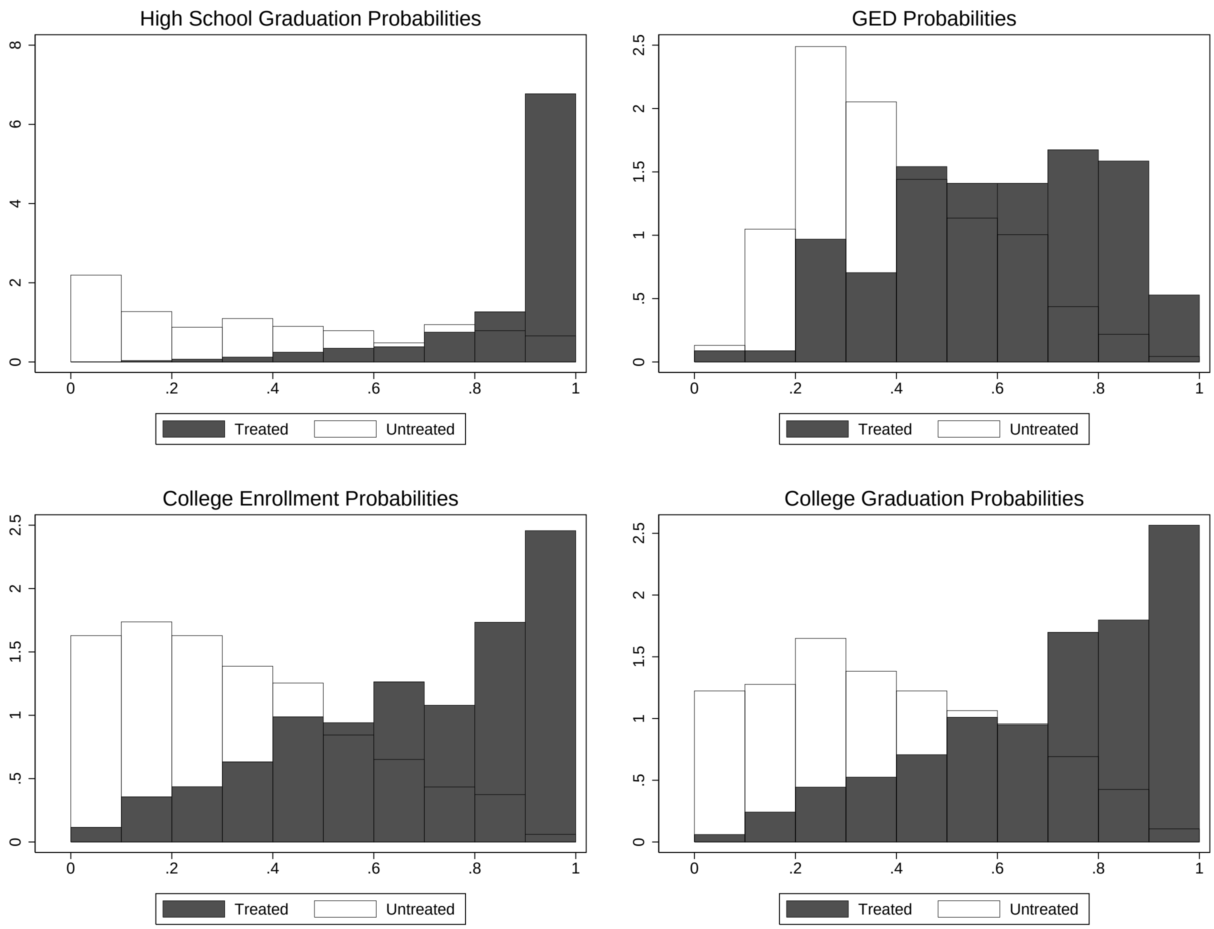
We use two versions of nearest-neighbor matching. These are reported in columns five and six. "NNM(3)-F" is nearest-neighbor matching using the three nearest neighbors measured by the Mahalanobis metric:

$$\|\boldsymbol{Z}_i - \boldsymbol{Z}_{k(i)}\| = \left(\boldsymbol{Z}_i - \boldsymbol{Z}_{k(i)}\right)' \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{\mathbf{-1}} \left(\boldsymbol{Z}_i - \boldsymbol{Z}_{k(i)}\right) \tag{A.14}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$ is the sample covariance of $\boldsymbol{Z}$, i is a subscript denoting the individual matched, and $k(i)$ is the candidate index of observations for matching to i. We use both treatments and controls to estimate ATE conditional on $Q_j = 1$. We form three matches for each observation in the sample $Q_j = 1$ for which $D_j = 0$ (using observations for which $D_j = 1$), and three matches for persons for which $D_j = 1$ (using observations for which $D_j = 0$). We sum these matches for each sample ($D_j = 0$ and $D_j = 1$), and then weight two resulting matching estimators, respectively, by the sample proportion with $D_j = 0$ and $D_j = 1$. We add the second weighted matching estimator (for $D_j = 1$) to the first weighted matching estimator (for $D_j = 0$) to produce an estimated ATE for $Q_j = 1$. We use all of the background and "exclusion restriction" variables of Table 1 to construct node-specific $\boldsymbol{Z}$. We match on the variables indicated at the base of Table 3: Bartlett cognitive and non-cognitive factors and an index of the $\boldsymbol{Z}$ variables constructed from the prediction from a node-specific linear probability model of D_j on $\boldsymbol{Z}_j$ conditional on $Q_j = 1$. The column labeled "PSM-F" reports estimates from a nearest-neighbor estimator formed using a propensity score estimator $Prob(D_j = 1|\boldsymbol{Z}_j, Q_j = 1)$ using a probit model. We use only a single nearest neighbor.

There is overlap in the support of the estimated propensity scores for the $D_j = 1$ and $D_j = 0$ samples for most nodes j except the node of high school graduation. See Figure A16, which graphs the density of the estimated transition probabilities $(Pr(Dj = 1|\boldsymbol{Z}, Q_j = 1))$ for individuals in our sample.

Figure A16: Supports of the Propensity Score at Each Decision Node



Notes: Each plot is for the population who reaches that decision node in the data. "Treated" are those who choose to complete the reported level of schooling, while "Untreated" are those who choose to not complete the reported level of schooling (but reach the decision node). Probabilities are estimated by a probit model that controls for the set of control variables and decision-specific instruments used and reported in the paper.

A.19 Comparing Model Parameter Estimates from the Full Sample and the Sample of White Males

This section compares our model to the model estimated only on white males. Each table shows the estimated parameters for our main specification and the parameters for the restricted model. Overall, we find that the estimated parameters are quite similar when controlling for race as we do in our main specification or when restricting the model only to white males.

Table A75: Model Parameter Estimates for Log Wages at Age 30 (comparing estimation on the full sample on white males)

Estimation on Full Sample:										
Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.200	0.068	-0.278	0.089	-0.229	0.052	-0.247	0.081	-0.066	0.079
Hispanic	-0.178	0.077	-0.050	0.143	-0.004	0.065	-0.136	0.103	-0.009	0.109
Broken Home	-0.022	0.047	0.139	0.073	0.064	0.038	-0.082	0.059	-0.063	0.056
Number of Sibs	-0.006	0.009	0.009	0.014	-0.007	0.007	0.008	0.011	0.006	0.010
Mother's HGC	0.004	0.011	0.006	0.017	-0.001	0.008	0.014	0.012	0.013	0.009
Father's HGC	0.008	0.009	0.019	0.012	0.012	0.006	-0.006	0.009	0.006	0.007
Age	0.151	0.214	0.331	0.329	-0.136	0.133	0.191	0.211	0.097	0.175
Age Squared	-0.004	0.006	-0.007	0.009	0.004	0.003	-0.005	0.005	-0.002	0.005
Fam. Income 1979	0.008	0.003	0.008	0.004	0.007	0.002	0.008	0.002	0.005	0.001
Constant	0.591	2.015	-1.690	3.079	3.251	1.255	0.451	1.982	1.361	1.650
Local Unemp.	-1.021	1.192	0.905	1.605	0.511	0.759	-1.520	1.199	-0.984	1.050
Northeast 30	0.291	0.073	0.125	0.115	0.070	0.041	0.128	0.070	0.183	0.050
South 30	0.078	0.058	0.024	0.089	-0.046	0.038	0.016	0.063	-0.025	0.049
West 30	0.060	0.076	0.005	0.107	0.046	0.045	0.062	0.066	-0.001	0.056
Urban 30	0.055	0.054	0.108	0.084	0.098	0.033	0.117	0.060	0.139	0.056
Factor 1	0.095	0.052	0.147	0.059	0.157	0.024	0.040	0.042	0.233	0.042
Factor 2	-0.056	0.054	0.076	0.073	-0.049	0.029	0.032	0.051	0.057	0.045
Estimation on White Males:										
	HSDropout		GED		HS Grad.		Some College		College Grad.	
Broken Home	-0.074	0.052	0.147	0.088	0.070	0.044	-0.081	0.070	-0.017	0.062
Number of Sibs	-0.009	0.011	0.020	0.019	-0.008	0.008	0.009	0.014	0.005	0.011
Mother's HGC	-0.005	0.013	-0.002	0.021	-0.001	0.010	0.014	0.015	0.017	0.010
Father's HGC	0.017	0.011	0.027	0.016	0.013	0.007	-0.007	0.010	0.007	0.007
Age	0.289	0.234	0.548	0.381	-0.202	0.142	0.058	0.227	0.096	0.181
Age Squared	-0.008	0.006	-0.013	0.010	0.006	0.004	-0.001	0.006	-0.002	0.005
Fam. Income 1979	0.010	0.003	0.011	0.005	0.006	0.002	0.008	0.002	0.004	0.001
Constant	-0.620	2.207	-3.823	3.579	3.924	1.346	1.720	2.135	1.258	1.705
Local Unemp	-1.158	1.259	0.537	1.852	0.606	0.799	-1.742	1.283	-0.179	1.124
Northeast 30	0.211	0.086	0.185	0.146	0.077	0.043	0.108	0.072	0.186	0.051
South 30	0.091	0.061	0.124	0.107	-0.033	0.041	0.036	0.067	-0.008	0.052
West 30	0.063	0.080	0.014	0.126	-0.005	0.047	0.048	0.070	-0.009	0.058
Urban 30	0.116	0.058	0.126	0.096	0.112	0.034	0.146	0.062	0.131	0.057
Factor 1	0.091	0.053	0.164	0.065	0.138	0.025	0.034	0.042	0.247	0.042
Factor 2	-0.062	0.057	0.024	0.089	-0.083	0.032	0.039	0.056	0.043	0.047

Notes: The top panel shows the parameter estimates for the model estimated with the full sample (see Equation (A.2) in Section A.4). In the bottom panel, the estimation is restricted to white males.

Table A76: Model Parameter Estimates for Log Present Value of Wages (comparing estimation on the full sample on white males)

Estimation on Full Sample:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.745	0.089	-0.351	0.114	-0.345	0.060	-0.318	0.091	-0.104	0.101
Hispanic	-0.152	0.113	-0.076	0.192	0.200	0.080	-0.315	0.109	0.315	0.137
Broken Home	-0.079	0.065	-0.044	0.095	-0.042	0.044	-0.122	0.066	-0.120	0.070
Number of Sibs	-0.037	0.012	0.031	0.018	-0.005	0.008	-0.003	0.014	-0.002	0.013
Mother's HGC	0.026	0.015	0.032	0.022	0.018	0.010	0.015	0.014	0.036	0.012
Father's HGC	0.016	0.012	0.021	0.016	0.019	0.007	-0.010	0.010	0.003	0.009
Age	-0.270	0.267	-0.300	0.379	0.012	0.143	0.025	0.221	-0.229	0.195
Age Squared	0.007	0.007	0.008	0.010	-0.000	0.004	-0.001	0.006	0.006	0.005
Fam. Income 1979	0.018	0.004	0.017	0.005	0.012	0.002	0.008	0.002	0.008	0.002
Constant	14.028	2.571	13.716	3.621	11.454	1.382	11.966	2.111	14.092	1.879
Urban 17	0.216	0.070	0.038	0.111	0.045	0.039	0.175	0.062	0.066	0.060
South 17	0.066	0.076	0.157	0.115	0.024	0.045	0.029	0.069	0.031	0.058
West 17	-0.144	0.101	-0.157	0.137	-0.027	0.052	0.116	0.076	-0.143	0.078
Northeast 17	0.162	0.097	-0.001	0.147	0.035	0.048	0.101	0.076	0.158	0.058
Factor 1	0.429	0.072	0.408	0.075	0.180	0.029	0.079	0.046	0.230	0.053
Factor 2	0.020	0.076	0.017	0.093	-0.089	0.035	-0.010	0.056	0.138	0.057

Estimation on White Males:

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Broken Home	-0.086	0.031	-0.044	0.090	0.021	0.049	-0.142	0.079	-0.061	0.079
Number of Sibs	-0.019	0.006	-0.018	0.020	-0.002	0.009	-0.001	0.017	-0.008	0.014
Mother's HGC	0.029	0.006	0.008	0.022	0.016	0.011	0.015	0.017	0.044	0.013
Father's HGC	0.014	0.005	0.023	0.016	0.016	0.008	-0.019	0.011	0.002	0.009
Age	-0.085	0.100	-0.021	0.352	0.042	0.149	-0.090	0.246	-0.215	0.202
Age Squared	0.002	0.003	0.001	0.009	-0.001	0.004	0.002	0.006	0.005	0.005
Fam. Income 1979	0.010	0.001	0.016	0.005	0.010	0.002	0.007	0.002	0.007	0.002
Constant	12.408	0.965	11.491	3.368	11.205	1.441	13.144	2.350	13.894	1.946
Urban 17	0.114	0.028	0.046	0.100	0.104	0.040	0.215	0.067	0.069	0.062
South 17	0.032	0.030	0.213	0.111	0.012	0.046	0.085	0.076	0.023	0.061
West 17	-0.068	0.036	-0.174	0.128	-0.100	0.053	0.140	0.083	-0.126	0.084
Northeast 17	0.128	0.033	0.265	0.145	0.033	0.049	0.116	0.079	0.132	0.060
Factor 1	0.238	0.018	0.388	0.065	0.145	0.027	0.031	0.048	0.250	0.055
Factor 2	0.073	0.023	0.003	0.086	-0.094	0.037	0.028	0.063	0.096	0.061

Notes: The top panel shows the parameter estimates for the model estimated with the full sample (see Equation (A.2) in Section A.4). In the bottom panel, the estimation is restricted to white males.

Table A77: Model Parameter Estimates for Smoking at Age 30 (comparing estimation on the full sample on white males)

Estimation on Full Sample:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	-0.199	0.277	-0.106	0.271	0.216	0.180	-0.349	0.288	-0.081	0.359
Hispanic	-0.943	0.353	-1.298	0.467	-0.296	0.245	-0.413	0.342	0.520	0.431
Broken Home	0.317	0.209	-0.225	0.222	0.037	0.129	0.623	0.199	-0.008	0.253
Number of Sibs	0.104	0.043	-0.020	0.045	-0.003	0.025	0.053	0.038	-0.019	0.043
Mother's HGC	0.004	0.046	-0.072	0.054	0.018	0.028	0.031	0.042	-0.076	0.040
Father's HGC	0.054	0.040	0.031	0.038	0.046	0.022	0.008	0.029	0.076	0.032
Age	0.240	0.822	0.393	0.910	-0.343	0.417	0.543	0.675	-0.297	0.685
Age Squared	-0.004	0.021	-0.008	0.024	0.009	0.011	-0.014	0.017	0.009	0.018
Fam. Income 1979	0.013	0.013	-0.006	0.012	-0.009	0.005	0.004	0.007	-0.003	0.005
Constant	-4.229	7.854	-3.494	8.633	2.423	4.018	-6.544	6.445	1.353	6.608
Northeast 30	0.073	0.327	-0.323	0.361	0.132	0.139	-0.281	0.239	-0.199	0.215
South 30	0.023	0.246	-0.432	0.281	0.119	0.129	-0.120	0.208	0.010	0.198
West 30	-0.291	0.303	-0.062	0.330	-0.212	0.155	0.018	0.217	-0.425	0.267
Urban 30	-0.104	0.225	0.147	0.257	0.043	0.113	0.112	0.196	0.421	0.286
Factor 1	-0.341	0.229	-0.385	0.183	-0.023	0.083	-0.068	0.135	-0.233	0.181
Factor 2	-0.444	0.233	-0.166	0.215	-0.131	0.103	-0.153	0.164	-0.361	0.195

Estimation with White Males

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
Broken Home	0.589	0.270	-0.155	0.254	0.049	0.148	0.517	0.235	-0.062	0.290
Number of Sibs	0.132	0.061	0.014	0.057	0.008	0.029	0.064	0.045	0.003	0.045
Mother's HGC	0.019	0.064	-0.098	0.061	-0.007	0.033	0.028	0.051	-0.074	0.043
Father's HGC	0.076	0.049	0.051	0.045	0.051	0.024	0.022	0.032	0.069	0.034
Age	-0.013	0.985	-0.126	1.033	-0.418	0.455	0.593	0.728	-0.260	0.715
Age Squared	0.003	0.026	0.006	0.027	0.010	0.012	-0.015	0.019	0.008	0.018
Fam. Income 1979	0.006	0.015	-0.004	0.014	-0.008	0.006	0.001	0.007	-0.000	0.005
Constant	-2.260	9.389	1.275	9.763	3.358	4.383	-6.862	6.943	0.891	6.901
Northeast 30	0.181	0.434	-0.436	0.426	0.122	0.148	-0.230	0.247	-0.206	0.224
South 30	-0.018	0.278	-0.538	0.323	0.125	0.139	-0.118	0.222	0.013	0.211
West 30	-0.194	0.357	-0.062	0.385	-0.112	0.163	-0.041	0.234	-0.393	0.281
Urban 30	-0.113	0.270	0.185	0.278	0.101	0.119	0.051	0.204	0.391	0.286
Factor 1	-0.183	0.266	-0.429	0.191	-0.032	0.084	-0.003	0.139	-0.077	0.189
Factor 2	-0.634	0.290	-0.079	0.236	-0.142	0.114	-0.310	0.183	-0.372	0.211

Notes: The top panel shows the parameter estimates for the model estimated with the full sample (see Equation (A.2) in Section A.4). In the bottom panel, the estimation is restricted to white males.

Table A78: Model Parameter Estimates for Health Limits Work (comparing estimation on the full sample on white males)

Estimation on Full Sample:

Variables	HS Dropout		GED		HS Grad.		Some College		College Grad.	
	β	Std Err	β	Std Err	β	Std Err	β	Std Err	β	Std Err
Black	0.360	0.228	0.033	0.242	0.102	0.174	0.173	0.269	0.142	0.321
Hispanic	-0.178	0.292	-0.352	0.391	0.129	0.214	0.204	0.329	-0.782	0.612
Broken Home	0.079	0.173	-0.061	0.199	-0.051	0.132	0.303	0.195	0.250	0.238
Number of Sibs	0.002	0.033	-0.032	0.039	-0.001	0.025	0.038	0.038	0.019	0.044
Mother's HGC	-0.009	0.040	-0.066	0.048	-0.016	0.027	-0.021	0.042	-0.015	0.043
Father's HGC	0.011	0.033	-0.033	0.034	-0.051	0.021	0.023	0.030	-0.003	0.031
Age	0.034	0.722	-0.388	0.803	0.869	0.436	0.205	0.675	-0.349	0.725
Age Squared	0.002	0.019	0.012	0.021	-0.021	0.011	-0.006	0.017	0.012	0.018
Fam. Income 1979	-0.012	0.011	-0.013	0.010	-0.000	0.006	0.003	0.007	0.004	0.005
Constant	-1.774	6.969	4.121	7.706	-8.930	4.214	-2.751	6.442	0.929	7.047
Urban 17	-0.165	0.191	-0.002	0.237	0.204	0.120	-0.280	0.187	-0.113	0.224
South 17	0.137	0.206	-0.288	0.242	-0.054	0.136	-0.195	0.207	0.160	0.216
West 17	0.407	0.277	0.060	0.275	0.249	0.153	0.069	0.224	0.407	0.270
Northeast 17	0.374	0.259	-0.219	0.305	0.034	0.140	-0.418	0.242	0.067	0.224
Factor 1	-0.312	0.190	-0.269	0.159	-0.309	0.085	0.013	0.140	-0.347	0.175
Factor 2	0.337	0.201	-0.169	0.195	-0.146	0.100	-0.147	0.167	-0.124	0.203

Estimation with White Males

Variables	HSDropout		GED		HS Grad.		Some College		College Grad.	
Broken Home	-0.078	0.208	-0.189	0.228	-0.069	0.159	0.205	0.234	0.099	0.273
Number of Sibs	0.067	0.043	-0.037	0.050	-0.015	0.030	0.057	0.046	0.058	0.047
Mother's HGC	-0.047	0.054	-0.037	0.057	-0.033	0.034	-0.003	0.051	-0.038	0.046
Father's HGC	0.039	0.040	-0.069	0.042	-0.055	0.025	0.025	0.034	-0.003	0.032
Age	0.560	0.859	-0.532	0.902	0.919	0.489	-0.292	0.737	-0.085	0.762
Age Squared	-0.010	0.022	0.015	0.023	-0.022	0.013	0.008	0.019	0.005	0.019
Fam. Income 1979	-0.004	0.012	-0.008	0.012	0.003	0.006	0.001	0.007	0.005	0.005
Constant	-7.810	8.334	5.948	8.667	-9.222	4.733	1.582	7.026	-1.450	7.423
Urban 17	-0.064	0.213	-0.173	0.256	0.189	0.130	-0.194	0.205	-0.063	0.235
South 17	0.234	0.234	-0.391	0.282	-0.063	0.151	-0.250	0.226	0.236	0.227
West 17	0.351	0.319	-0.108	0.314	0.232	0.166	-0.150	0.247	0.502	0.288
Northeast 17	0.489	0.328	-0.140	0.366	-0.062	0.154	-0.396	0.250	0.114	0.235
Factor 1	-0.441	0.220	-0.438	0.172	-0.249	0.087	-0.022	0.147	-0.389	0.173
Factor 2	0.169	0.224	0.066	0.219	-0.210	0.115	-0.141	0.189	0.016	0.217

Notes: The top panel shows the parameter estimates for the model estimated with the full sample (see Equation (A.2) in Section A.4). In the bottom panel, the estimation is restricted to white males.

A.20 Alternative Policy-Relevant Treatment Effects

In the paper, we considered a tuition subsidy as a policy experiment, but many other experiments can be considered and need not correspond to an instrument in the data. Consider a simulated instrument that increases the probability of graduating from high school by 5%. We can then estimate the impact of the increase and how many of those induced to graduate high school go on to complete additional education.

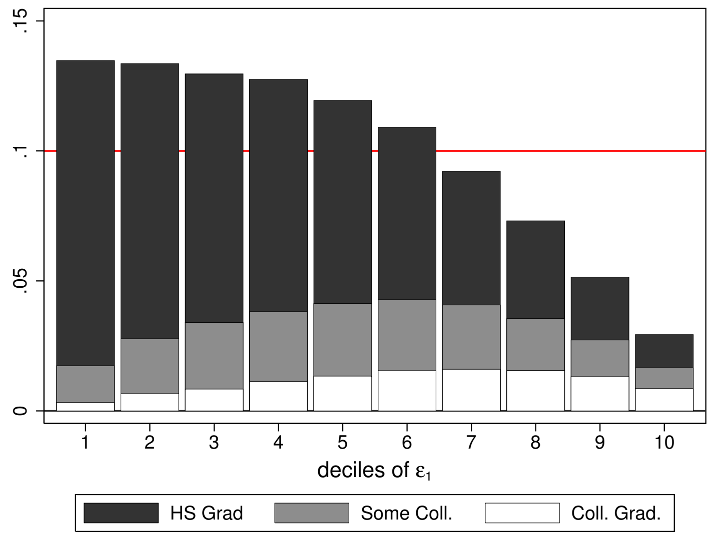
Figure A17 shows which individuals are induced to graduate from high school by decile of the unobserved heterogeneity acted on by the agent but unobserved by the econometrician. The figure further decomposes switchers into those that stop, those that go on to enroll in college, and those that go on to graduate from college. Overall, some individuals in each decile are induced to graduate, but the bulk of the movers are in the lower deciles. Individuals that move in higher deciles are more likely to then go on and earn additional education.

Table A79: PRTE: Increase in the Probability of Graduating from High School

	PRTE
Log Wages	0.093
PV Log Wages	0.209
Health Limits Work	-0.113
Smoking	-0.250

Notes: Table shows the Policy-Relevant Treatment Effect (PRTE) of reducing the probability of enrolling in college by 5%.

Figure A17: PRTE: Who Is Induced to Switch from a 5% Increase in the Probability of Graduating from High School?



Notes: The figure plots the proportion of individuals induced to switch from the policy that lay in each decile of ϵ_1, where is the unobserved component of the educational choice model. The bars are further decomposed into those that are induced to switch that stop at high school graduation, those that go on to enroll in college, and those that go on to graduate from college.

A.21 Evidence on Treatment Effects by Decile of Cognitive and Socio-Emotional Endowments and on the Direct Role of Those Endowments Fixing the Schooling Level

This section provides an overview of: (i) treatment effects by decile of cognitive and socio-emotional endowments conditional on schooling level; and (ii) the direct role of cognitive and socio-emotional endowments fixing the schooling level.

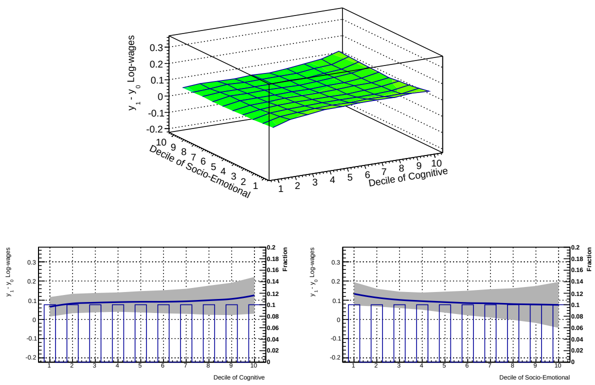
A.21.1 The Effect of Endowments on Treatment Effects

These figures complement the results from the treatment effect tables.

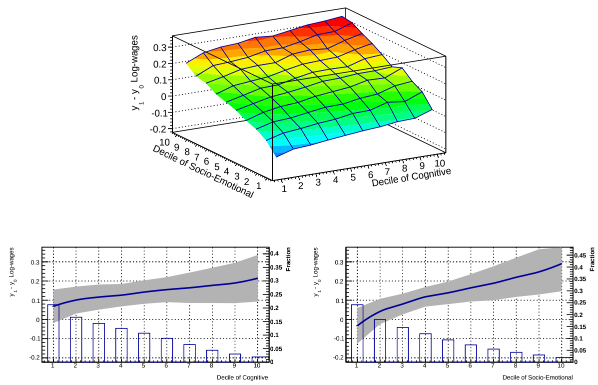
Each figure analyzes the average effects of education on the outcome of interest. For a particular outcome, the effect is defined as the difference in the outcomes associated with two schooling levels (not necessarily final or terminal schooling levels). For each pairwise comparison of outcomes, let Y_0 and Y_1 denote the outcomes associated with schooling levels 0 and 1, respectively. Importantly, each schooling level might provide the option to pursue higher schooling levels. Final schooling levels do not allow for further options. Notice that in the figures, final schooling levels are highlighted using bold letters. For each pair of schooling levels 0 and 1, the first figure (top) presents $E(Y_1 - Y_0|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments. $E(Y_1 - Y_0|d^C, d^{SE})$ is computed for those who reach the decision node involving a decision between levels 0 and 1. The second figure (bottom left) presents $E(Y_1 - Y_0|d^C)$ so that the socio-emotional 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 levels 0 and 1. The last figure (bottom right) presents $E(Y_1 - Y_0|d^{SE})$ as well as, for a given decile of socio-emotional endowment, the fraction of individuals visiting the node leading to the educational decision involving levels 0 and 1.

Figure A18: Average Treatment Effect of Education on Log Wages at Age 30, by *Decision Node* and Endowment Levels

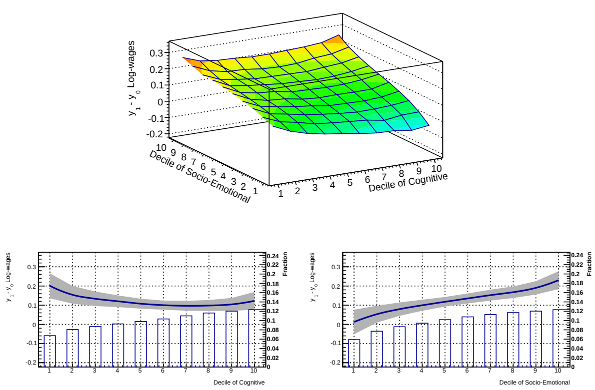
A. Graduating from HS vs. Dropping from HS



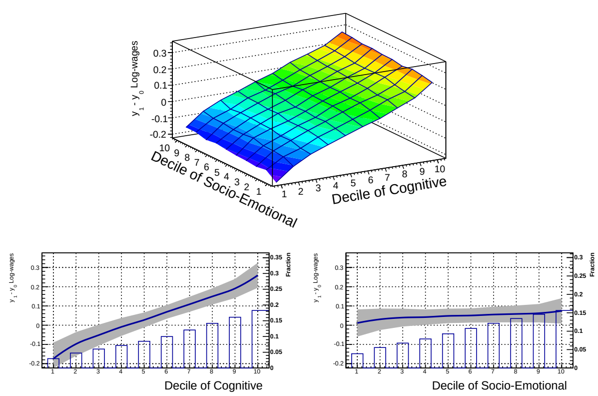
B. Getting a GED vs. **HS Dropout**



C. College Enrollment vs. **HS Graduate**

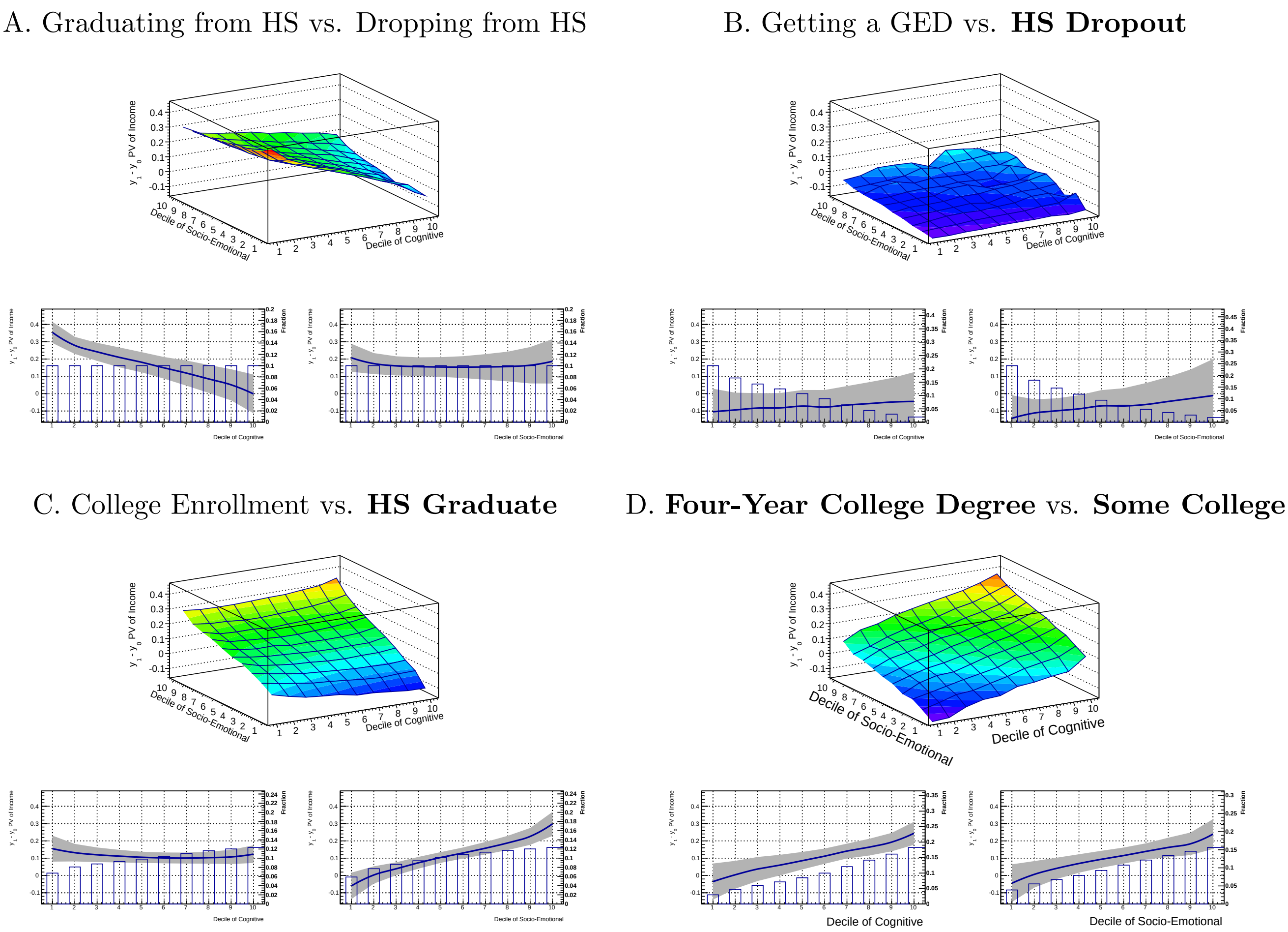


D. **Four-Year College Degree** vs. **Some College**



Notes: Each panel in this figure studies the average effect of an educational decision for those individuals visiting the decision node. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. For each educational decision node, the first figure (top) presents $ATE_j\left(\theta \in (d^C, d^{SE})\right)$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments for the full population. The second figure (bottom left) presents $ATE_j\left(\theta^C \in d^C\right)$ so that the socio-emotional factor is integrated out. The bars in this figure display the fraction of individuals visiting the node in each decile of cognitive endowment. The last figure (bottom right) presents $ATE_{j,j''}\left(\theta^{SE} \in d^{SE}\right)$ and the fraction of individuals visiting the node in a given decile of socio-emotional endowment.

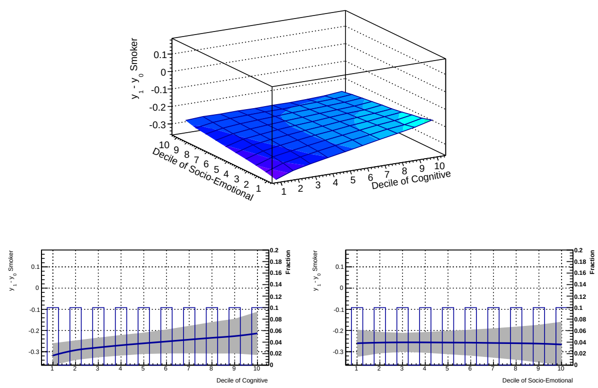
Figure A19: Average Treatment Effect of Education on Present Value of Wages, by *Decision Node* and Endowment Levels



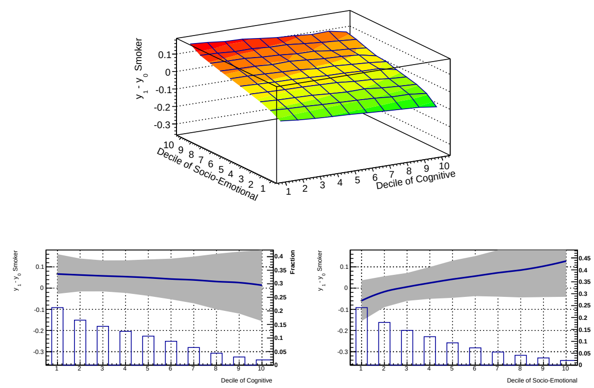
Notes: Each panel in this figure studies the average effect of an educational decision for those individuals visiting the decision node. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. For each educational decision node, the first figure (top) presents $ATE_j\left(\theta \in (d^C, d^{SE})\right)$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments for the full population. The second figure (bottom left) presents $ATE_j\left(\theta^C \in d^C\right)$ so that the socio-emotional factor is integrated out. The bars in this figure display the fraction of individuals visiting the node in each decile of cognitive endowment. The last figure (bottom right) presents $ATE_{j,j''}\left(\theta^{SE} \in d^{SE}\right)$ and the fraction of individuals visiting the node in a given decile of socio-emotional endowment.

Figure A20: Average Treatment Effect of Education on Smoking, by *Decision Node* and Endowment Levels

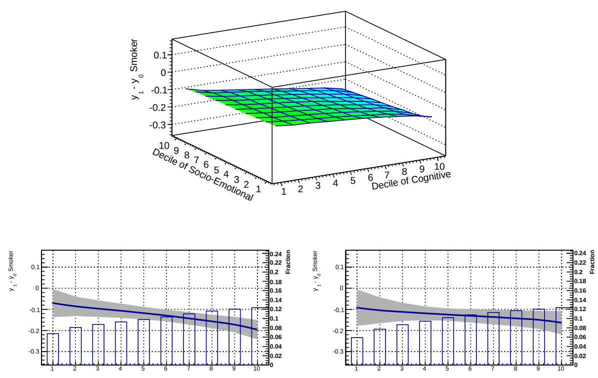
A. Graduating from HS vs. Dropping from HS



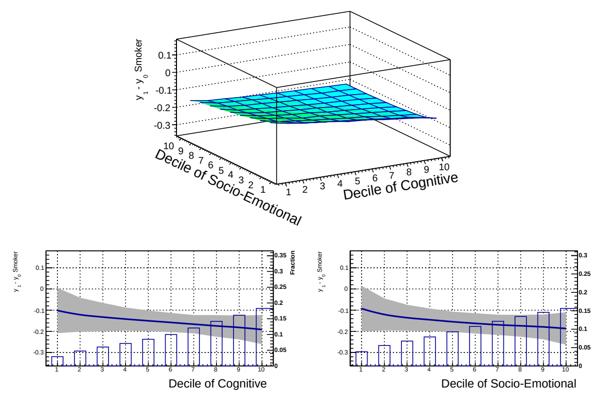
B. Getting a GED vs. **HS Dropout**



C. College Enrollment vs. **HS Graduate**



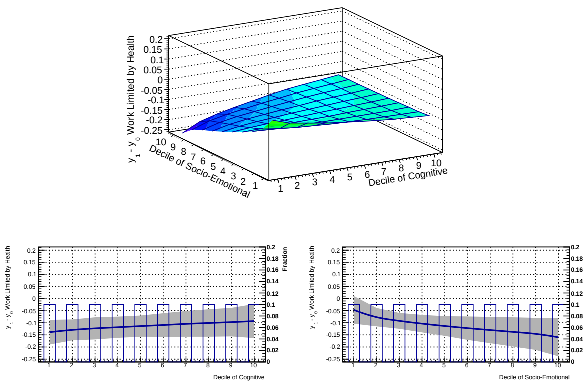
D. **Four-Year College Degree** vs. **Some College**



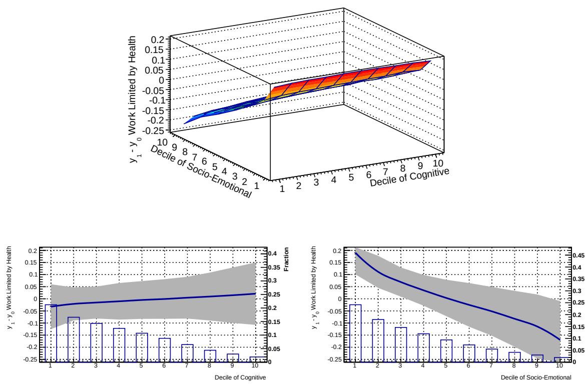
Notes: Each panel in this figure studies the average effect of an educational decision for those individuals visiting the decision node. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. For each educational decision node, the first figure (top) presents $ATE_j\left(\theta \in (d^C, d^{SE})\right)$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments for the full population. The second figure (bottom left) presents $ATE_j\left(\theta^C \in d^C\right)$ so that the socio-emotional factor is integrated out. The bars in this figure display the fraction of individuals visiting the node in each decile of cognitive endowment. The last figure (bottom right) presents $ATE_{j,j''}\left(\theta^{SE} \in d^{SE}\right)$ and the fraction of individuals visiting the node in a given decile of socio-emotional endowment.

Figure A21: Average Treatment Effect of Education on Health Limits Work, by *Decision Node* and Endowment Levels

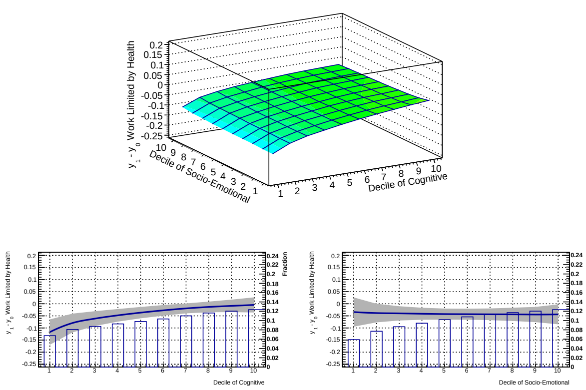
A. Graduating from HS vs. Dropping from HS



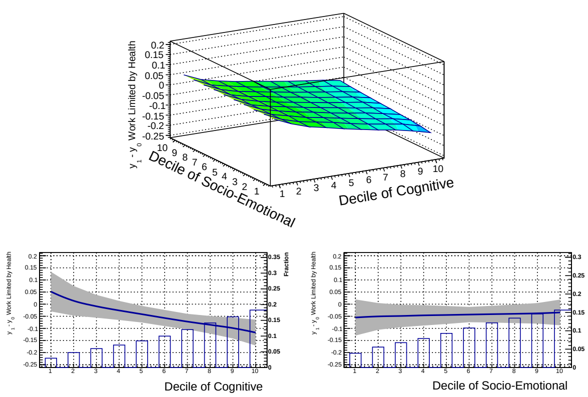
B. Getting a GED vs. **HS Dropout**



C. College Enrollment vs. **HS Graduate**



D. **Four-Year College Degree** vs. **Some College**



Notes: Each panel in this figure studies the average effect of an educational decision for those individuals visiting the decision node. Importantly, each schooling level might provide the option to pursue higher schooling levels, while final schooling levels do not provide an option. Final schooling levels are highlighted using bold letters. For each educational decision node, the first figure (top) presents $ATE_j\left(\theta \in (d^C, d^{SE})\right)$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments for the full population. The second figure (bottom left) presents $ATE_j\left(\theta^C \in d^C\right)$ so that the socio-emotional factor is integrated out. The bars in this figure display the fraction of individuals visiting the node in each decile of cognitive endowment. The last figure (bottom right) presents $ATE_{j,j''}\left(\theta^{SE} \in d^{SE}\right)$ and the fraction of individuals visiting the node in a given decile of socio-emotional endowment.

A.21.2 Testing if Endowments Affect Conditional Outcomes

Table A80 tests if the factor loadings in the education-specific outcomes are: (i) jointly equal to zero; and (ii) jointly equal to one another. The test is run separately for cognitive and socio-emotional factors. We find that we can always reject the null hypothesis that the cognitive loadings are jointly equal to zero. For the socio-emotional factor, we can reject the null for log PV wages and smoking. We generally cannot reject the null for either outcome that the factors are jointly equal to one another, except in the case of smoking for the socio-emotional factor.

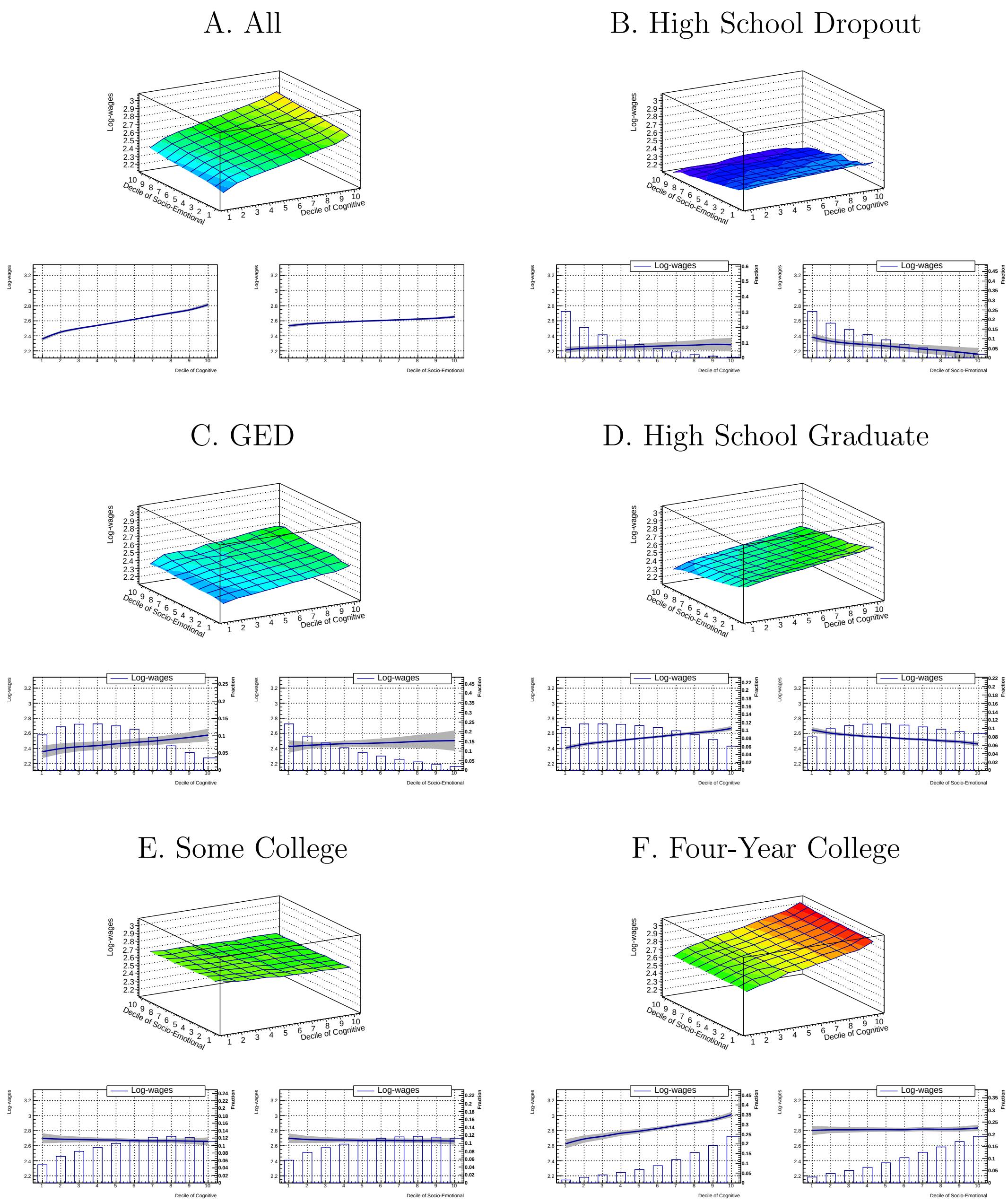
Table A80: Tests on The Estimated Factor Loadings on Cognitive and Socio-Emotional Factors by Outcome and Schooling Level

Variables	Tests	
	p-val$^{(a)}$	p-val$^{(b)}$
Log Wages		
Cognitive	0.000	0.213
Socio-Emotional	0.224	0.344
Log PV Wages		
Cognitive	0.000	0.464
Socio-Emotional	0.054	0.210
Smoking (Age 30)		
Cognitive	0.016	0.184
Socio-Emotional	0.015	0.053
Health Limits Work		
Cognitive	0.000	0.398
Socio-Emotional	0.202	0.397

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.

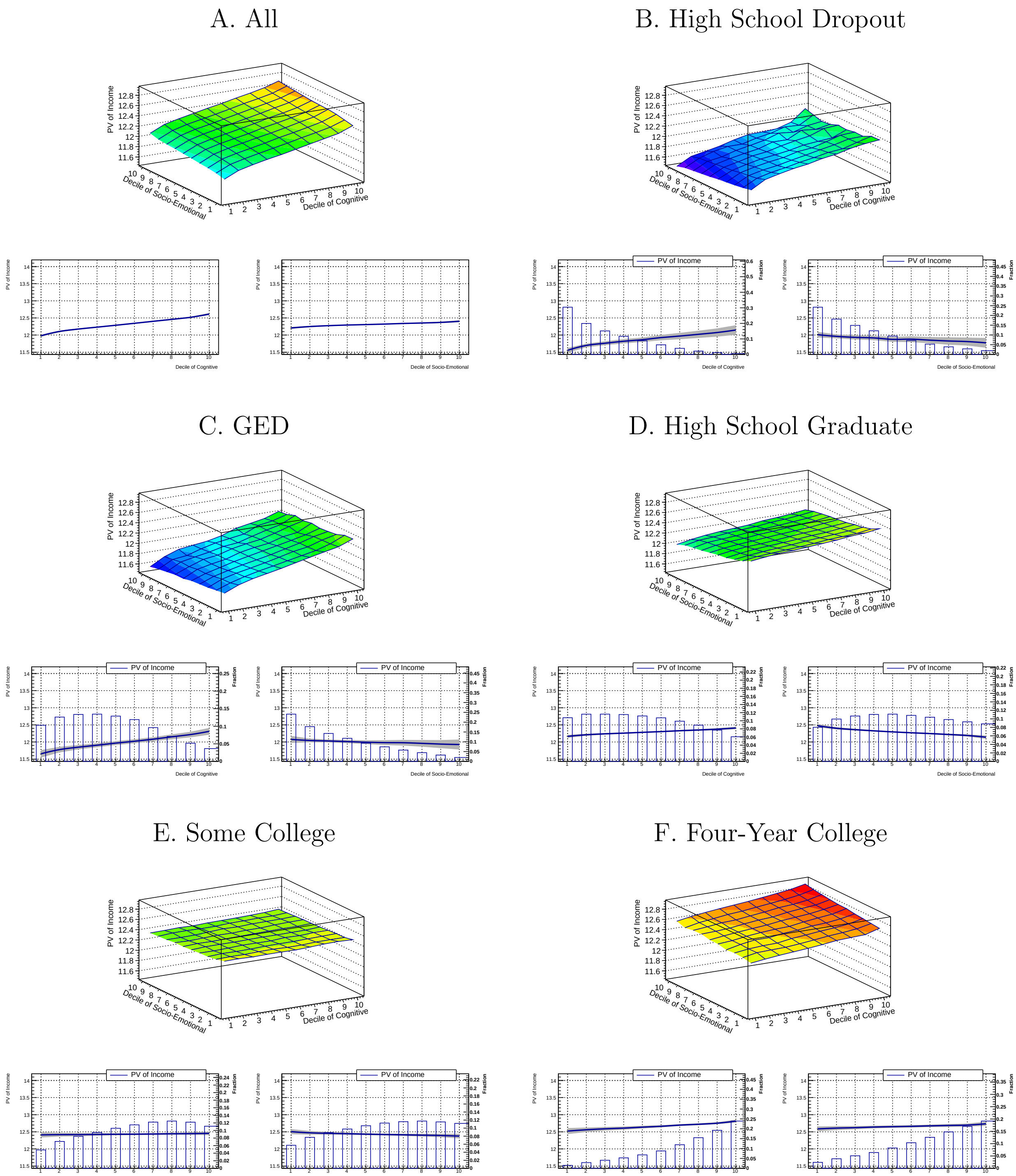
A.21.3 The Direct Impact of Endowments on Conditional Outcomes

Figure A22: The Effect of Cognitive and Socio-Emotional Endowments on Log Wages (age 30)



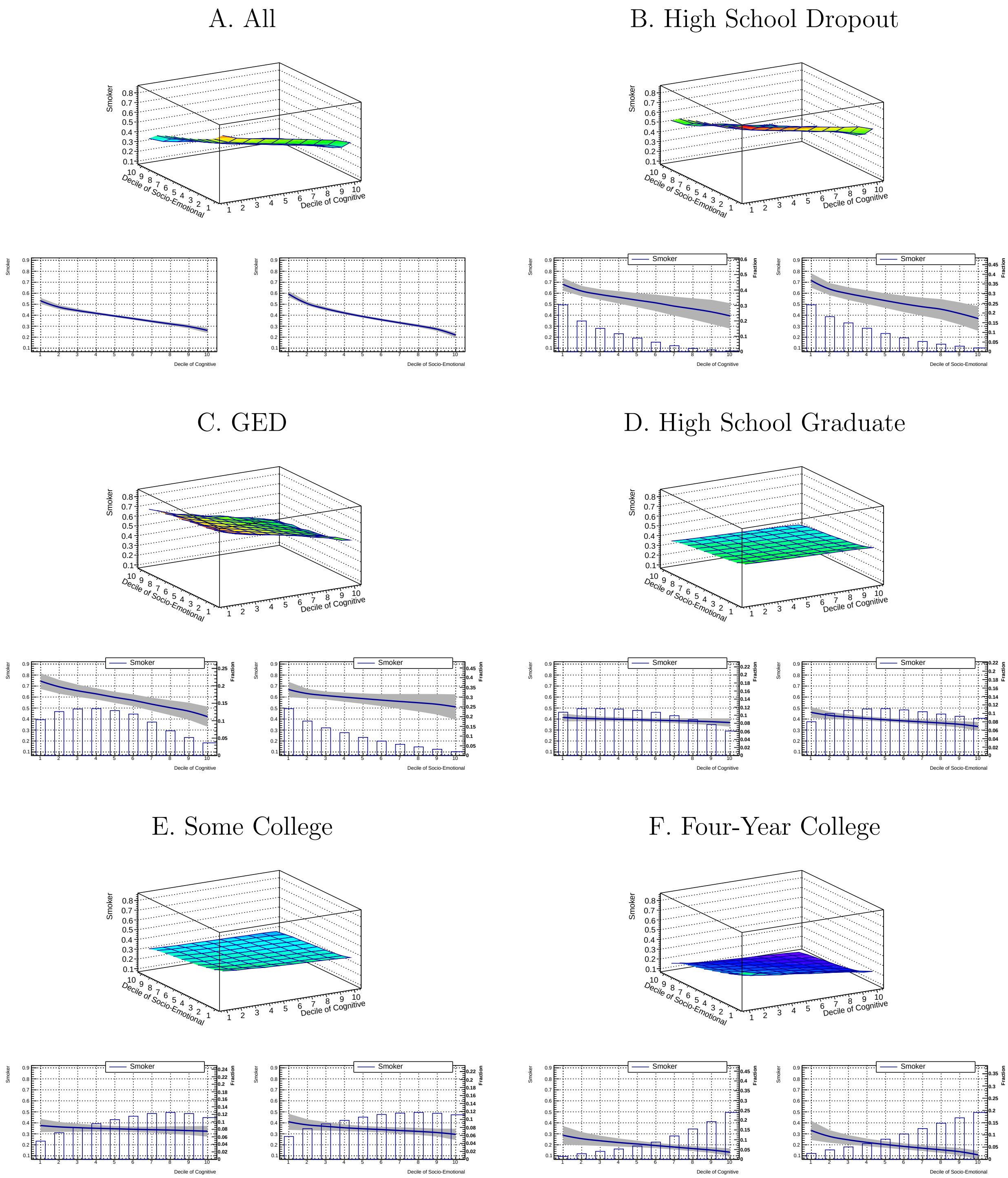
Notes: For each schooling level, we present three figures. The first figure (top) displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional endowments. Notice that we define "decile 1" as the decile with the lowest values of endowments and "decile 10" as the decile with the highest levels of endowments. The second figure (bottom left) displays the average levels of endowment across deciles of cognitive endowments. The bars in this figure indicate the fraction of individuals reporting the respective schooling level for each decile of cognitive endowment. The last figure (bottom right) mimics the structure of the second one but now for the socio-emotional endowment.

Figure A23: The Effect of Cognitive and Socio-Emotional Endowments on Log Present Value Wages



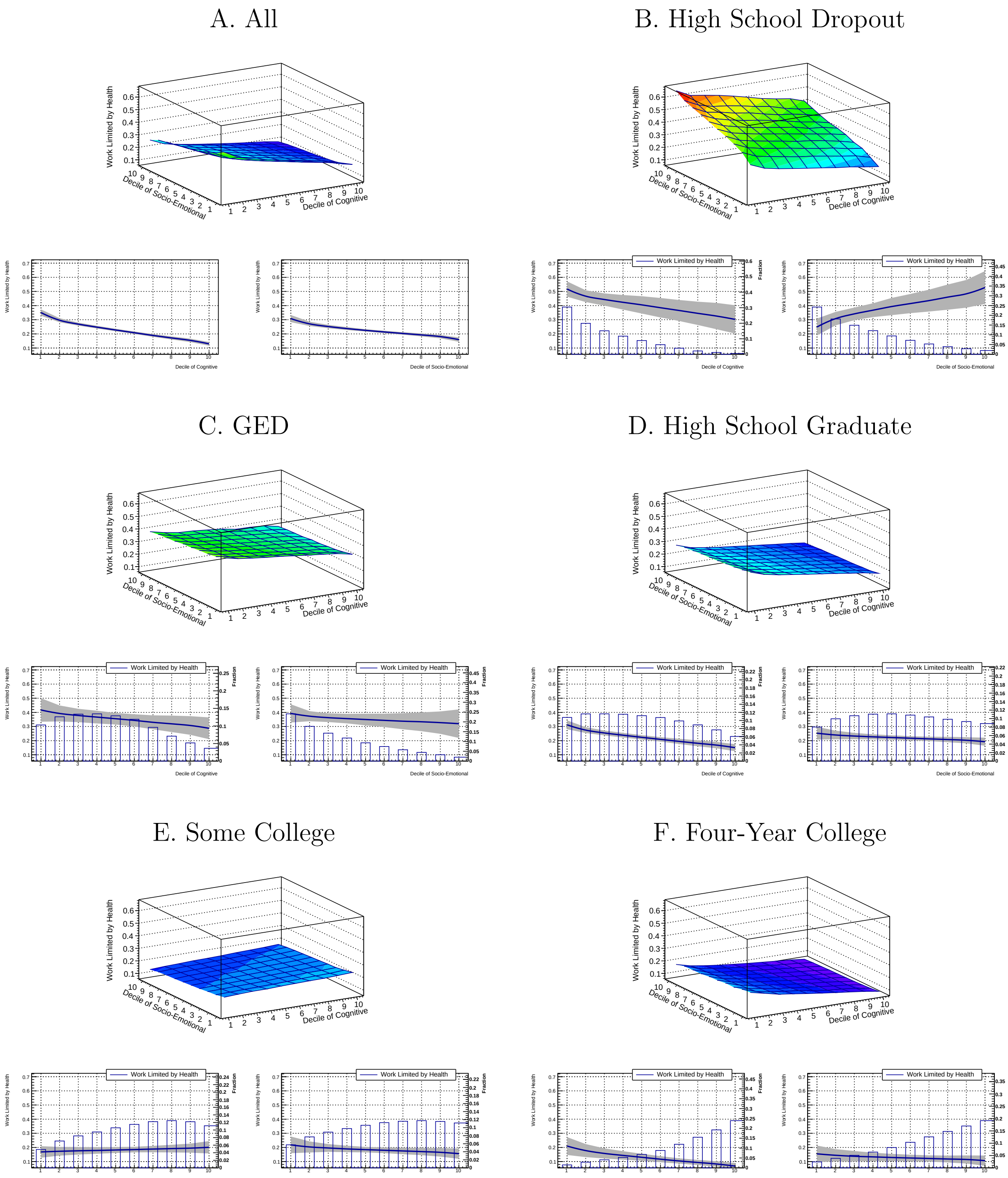
Notes: For each schooling level, we present three figures. The first figure (top) displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional endowments. Notice that we define "decile 1" as the decile with the lowest values of endowments and "decile 10" as the decile with the highest levels of endowments. The second figure (bottom left) displays the average levels of endowment across deciles of cognitive endowments. The bars in this figure indicate the fraction of individuals reporting the respective schooling level for each decile of cognitive endowment. The last figure (bottom right) mimics the structure of the second one but now for the socio-emotional endowment.

Figure A24: The Effect of Cognitive and Socio-Emotional Endowments on Smoking (age 30)



Notes: For each schooling level, we present three figures. The first figure (top) displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional endowments. Notice that we define "decile 1" as the decile with the lowest values of endowments and "decile 10" as the decile with the highest levels of endowments. The second figure (bottom left) displays the average levels of endowment across deciles of cognitive endowments. The bars in this figure indicate the fraction of individuals reporting the respective schooling level for each decile of cognitive endowment. The last figure (bottom right) mimics the structure of the second one but now for the socio-emotional endowment.

Figure A25: The Effect of Cognitive and Socio-Emotional Endowments on Health Limits Work



Notes: For each schooling level, we present three figures. The first figure (top) displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional endowments. Notice that we define “decile 1” as the decile with the lowest values of endowments and “decile 10” as the decile with the highest levels of endowments. The second figure (bottom left) displays the average levels of endowment across deciles of cognitive endowments. The bars in this figure indicate the fraction of individuals reporting the respective schooling level for each decile of cognitive endowment. The last figure (bottom right) mimics the structure of the second one but now for the socio-emotional endowment.

References

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

Albouy, Valerie, and Laurent Lequien, 2009. "Does Compulsory Education Lower Mortality?" *Journal of Health Economics* 28 (1):155–168.

Almlund, Mathilde, Angela Duckworth, James J. Heckman, and Tim Kautz, 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, vol. 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Wößmann, chap. 1. Amsterdam: Elsevier, pp. 1–181.

Anderson, Theodore W., and Herman Rubin, 1956. "Statistical Inference in Factor Analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5, edited by Jerzy Neyman. Berkeley, CA: University of California Press, pp. 111–150.

Arcidiacono, Peter, and Paul B. Ellickson, 2011. "Practical Methods for Estimation of Dynamic Discrete Choice Models." *Annual Review of Economics* 3 (1):363–394.

Arendt, Jacob N., 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification." *Economics of Education Review* 24 (2):149–160.

Auld, M. Christopher, and Nirmal Sidhu, 2005. "Schooling, Cognitive Ability and Health." *Health Economics* 14 (10):1019–1034.

Bartlett, Maurice S., 1937. "The Statistical Conception of Mental Factors." *British Journal of Psychology* 28 (1):97–104.

———, 1938. "Methods of Estimating Mental Factors." *Nature* 141:609–610.

Becker, Gary S., 1967. "Human Capital and the Personal Distribution of Income: An

Analytical Approach." Woytinsky Lecture no. 1. Ann Arbor: University of Michigan, Institute of Public Administration.

———, 1991. *A Treatise on the Family.* Cambridge, MA: Harvard University Press, enlarged edn.

Belzil, Christian, and Jorgen Hansen, 2002. "Unobserved Ability and the Return to Schooling." *Econometrica* 70 (5):2075–2091.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries, 2011. "Identification Problems in Personality Psychology." *Personality and Individual Differences* 51 (3: Special Issue on Personality and Economics):315–320.

Cameron, Stephen V., and James 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):455–499.

Card, David, 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, vol. 3A, edited by Orley C. Ashenfelter, and David Card, chap. 30. Amsterdam: Elsevier Science B.V., pp. 1801–1863.

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

Chen, Xiaohong, 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman, and Edward E. Leamer, chap. 76. Amsterdam: Elsevier Science B. V., pp. 4777–5752.

Conard, Maureen A., 2006. "Aptitude Is Not Enough: How Personality and Behavior Predict Academic Performance." *Journal of Research in Personality* 40 (3):339–346.

Credé, Marcus, and Nathan R. Kuncel, 2008. "Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance." *Prospectives on Psychological Science* 3 (6):425–453.

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

de Walque, Damien, 2007. "Does Education Affect Smoking Behaviors? Evidence Using the Vietnam Draft as an Instrument for College Education." *Journal of Health Economics* 26 (5):877–895.

Duckworth, Angela L., Patrick D. Quinn, and Eli Tsukayama, 2012. "What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades." *Journal of Educational Psychology* 104 (2):439–451.

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

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

Eisenhauer, Philipp, James J. Heckman, and Stefano Mosso, 2015. "Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments." *International Economic Review* 56 (2):331–357.

Farsides, Tom, and Ruth Woodfield, 2003. "Individual Differences and Undergraduate Academic Success: The Roles of Personality, Intelligence, and Application." *Personality and Individual Differences* 34 (7):1225–1243.

Geweke, John, and Michael Keane, 2001. "Computationally Intensive Methods for Integration in Econometrics." In *Handbook of Econometrics*, vol. 5, edited by James J. Heckman, and Edward E. Leamer, chap. 56. Amsterdam: Elsevier Science B. V., pp. 3463–3568.

Gladden, Tricia, and Christopher Taber, 2000. "Wage Progression among Less Skilled Workers." In *Finding Jobs: Work and Welfare Reform*, edited by David E. Card, and Rebecca M. Blank. New York: Russell Sage Foundation, pp. 160–192.

Grimard, Franque, and Daniel Parent, 2007. "Education and Smoking: Were Vietnam War Draft Avoiders Also More Likely to Avoid Smoking?" *Journal of Health Economics* 26 (5):896–926.

Gullone, Eleonora, and Susan Moore, 2000. "Adolescent Risk-Taking and the Five-Factor Model of Personality." *Journal of Adolescence* 23 (4):393–407.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen, 2004. "The Effect of Schooling and Ability on Achievement Test Scores." *Journal of Econometrics* 121 (1–2):39–98.

Heckman, James J., 1981. "The Empirical Content of Alternative Models of Labor Earnings." Unpublished manuscript, University of Chicago, Department of Economics.

———, 2001. "Econometrics and Empirical Economics." *Journal of Econometrics* 100 (1):3–6.

———, 2008. "The Principles Underlying Evaluation Estimators with an Application to Matching." *Annales d'Economie et de Statistiques* 91–92:9–73.

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

Heckman, James J., and Lakshmi K. Raut, 2013. "Intergenerational Long Term Effects of Preschool – Structural Estimates From a Discrete Dynamic Programming Model." Working Paper 19077, National Bureau of Economic Research.

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

Heckman, James J., and Edward J. Vytlacil, 2007. "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*, vol. 6B, edited by James J. Heckman, and Edward E. Leamer, chap. 71. Amsterdam: Elsevier Science B. V., pp. 4875–5143.

Herrnstein, Richard J., and Charles A. Murray, 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Horn, John L., 1965. "A Rationale and Test for the Number of Factors in Factor Analysis." *Psychometrika* 30 (2):179–185.

John, Oliver P., Richard W. Robins, and Lawrence A. Pervin, editors, 2008. *Handbook of Personality: Theory and Research*. New York, NY: Guilford Press, 3rd edn.

Judd, Kenneth L., 1998. *Numerical Methods in Economics*. Cambridge, MA: MIT Press.

Kane, Thomas J., and Cecilia E. Rouse, 1995. "Labor-Market Returns to Two- and Four-Year College." *American Economic Review* 85 (3):600–614.

Keane, Michael P., 2002. "Financial Aid, Borrowing Constraints, and College Attendance: Evidence from Structural Estimates." *The American Economic Review* 92 (2):293–297.

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

———, 2000. "Eliminating Race Differences in School Attainment and Labor Market Success." *Journal of Labor Economics* 18 (4):614–652.

———, 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review* 42 (4):1051–1103.

Kling, Jeffrey R., 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business and Economic Statistics* 19 (3):358–364.

Koning, Pierre, Dinand Webbink, and Nicholas G. Martin, 2015. "The Effect of Education on Smoking Behavior: New Evidence from Smoking Durations of a Sample of Twins." *Empirical Economics* 48 (4):1479–1497.

Noftle, Erik E., and Richard W. Robins, 2007. "Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores." *Personality Processes and Individual Differences* 93 (1):116–139.

Oreopoulos, Phillip, 2006. "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter." *American Economic Review* 96 (1):152–175.

Sander, William, 1998. "The Effects of Schooling and Cognitive Ability on Smoking and Marijuana Use by Young Adults." *Economics of Education Review* 17 (3):317–324.

Sullivan, Paul, 2010. "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search." *International Economic Review* 51 (1):289–317.

Williams, Benjamin, 2011. "Notes on Identification of Factor Models." Unpublished manuscript, University of Chicago, Department of Economics.