Determinants of Risk Behaviors in Early Adulthood: Interplay of Cognitive Ability, Noncognitive Skills, and Health

James Heckman, Rodrigo Pinto, Hongbo Wang [1]

University of Chicago

October 10, 2008

Abstract

This article explores the impact of human abilities in explaining a variety of economic and behavioral outcomes focusing Risk Behavior. We develop a model where human abilities are decomposed into three main skills: cognitive, noncognitive and health. We improve on the methodology developed by Hansen et al. (2004), Carneiro et al. (2003), and Heckman et al. (2006), by allowing skill interaction, which then rationalizes data on many outcomes of human life. We focus on the interaction of the estimated latent skills and risk behavior.

We found that cognitive, noncognitive skills have a major impact on schooling choices. We estimate the skills conditional on four schooling categories: college education, high school graduation, GED, and dropouts. We found the higher schooling level, the higher cognitive skills, as expected. One interesting result is that the noncognitive ability does not follow the same pattern. In our estimations, the GED recipients are the ones with the lowest levels of noncognitive ability. This result corroborates the analysis of Heckman et al. (2006), which state that the GEDs have lower cognitive ability compared to their average cognitive ability.

We also show that schooling, driven by cognitive and noncognitive skills, plays an important role in determining risky behaviors. As schooling levels increase, the probability of becoming a gang member or having an irresponsible sexual behavior decreases. In special, the age of the first intercourse increases with the schooling level. The decomposition of the effect of the skills into each outcome shows often that noncognitive ability plays a major part on determining the levels of many outcomes, in special, the risky behavior ones.

Keywords: Health economics, Bayesian Latent Factor, Risk Behavior.

JEL code: I21, C93.

1 Introduction

Recent research on the economics, psychology and biology of human development has analyzed the effects of the cognitive skills, noncognitive skills and health status on a variety of human outcomes. Examples of such works are Cunha and Heckman (2007, 2008), Cunha, Heckman, Lochner, and Masterov (2006), and Heckman (2007, 2008).

The present study examines the roles of genes in affecting the formation of a complex set of abilities and health statuses, which in turn determine a multitude of later social, economic, and health outcomes. Recent economics literature has recognized the importance of non-cognitive traits or personality, coupled with cognitive abilities, for multiple aspects of social, economic, and health outcomes (e.g. Borghans et al. (2007); Ryff and Singer (2005) and Heckman et al. (2006)). Health status has long been shown to interplay with socioeconomic status. Thus, a better understanding of group disparities in socioeconomic and health well-being entails studying the production, development, and synergy of cognition, non-cognitive capabilities/personality, and health status from a life-cyle perspective (i.e. Cunha and Heckman (2006) and Heckman (2006)).

We use data from the National Longitudinal Study of Adolescent Health (Add Health) to explore the impact of human abilities in explaining a variety of economic and behavioral outcomes focusing Risk Behavior. We develop a model where human abilities are decomposed into three main skills: cognitive, non-cognitive and health.

This paper contributes to the literature by examining the interaction of latent cognitive, non-cognitive skills and health with Risk behavior;

This paper proceeds as follows. Section 2 describes the Add Health data set. Section 3 presents our methodology. The main empirical analysis is presented in Section 4 while section 5 concludes. Supplementary material is provided in Web Appendices.[1]

2 Data

We draw on data from the National Longitudinal Study of Adolescent Health (Add Health), which allow us to examine risky behaviors during early adulthood of a nationally representative sample

---

[1] http://jenni.uchicago.edu/AddHealthRisk

of adolescents from the mid-90s. The Add Health study starts in 1994-1995[2] and has thus far completed two follow-up surveys in 1996 and 2001-2, respectively. It puts a strong focus on how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents' health and risk behaviors. [3]

In addition to demographic characteristics and the aforementioned contextual factors, the Add Health encompasses a wide variety of topics on adolescent development, such as academic performance, physical development, health, sexual attitude and behavior, romantic partnerships, psychological well-being, and behavioral problems. In particular, the three waves of surveys provide rich information on various risky behaviors (See below). Measures of physical health, including self-report of health status, unhealthy conditions, and morbidity, are numerous. The Add Health surveys also field questions that capture such socio-emotional traits as self-esteem, impulsivity, conscientiousness, and openness to experience. Finally, an adapted version of Peabody Vocabulary Test is administered at both Waves I and II to assess participants' verbal intelligence. Other standard cognitive assessments (e.g. ASVAB) are not available. Yet, the Wave III educational module, named Add Health and Academic Achievement (AHAA), makes a special effort to collect extensive information on respondents' academic profiles, including annual course sequences/performance in high school, which compensates to some extent for insufficient cognitive measures in the main surveys.

The Core Subsample   *Description* The baseline survey (Wave I) of Add Health, which is completed in 1995, randomly selects students from within 80 U.S. high schools, alongside their feeder schools. The current study only uses its core, nationally representative sample consisting of approximately 12,000 7th to 12th graders from Wave I (Bearman, Jones, and Udry 1997).[4] Those enrolled in grades 7 to 11 at Wave I were also re-interviewed one year later (Wave II).[5] The latest follow-up survey (Wave III), which takes place in 2001 and 2002, attempts to relocate all initial Wave I participants and succeeds in reinterviewing three quarters of them, including over 9000 respondents from the core sample.

---

[2] The response rate is 78.9 percent.

[3] In light of its focus on social contexts, the Wave I survey also interviews parents and school administrators and collected school/neighborhood information.

[4] The Add Health study also over-sampled a handful of selected populations, including blacks from well-educated families (at least one parent has a college degree), as well as Chinese, Cuban, and Puerto Rican adolescents.

[5] Most of the Wave I 12th graders are excluded from this follow-up.

Risky Behavior Outcomes   *Description* Add Health surveys collect rich information on respondents' development and well-being. This study focuses on risky behaviors that affects one's physical, mental, and behavioral health. We draw on measures from the latest interview at which all respondents is in their young adulthood. We focus on substance use (smoking, drug use, and drinking), sex-related behaviors (teenage sex, promiscuity and unprotected sex), unhealthy habits (skipping breakfast, lack of exercise, over-eating, and excessive fast food), and delinquency (violence, non-violent delinquency, gang membership, driving while drunk, and involvement with judicial system). With respect to substance use, we rely on self-report of smoking, drug use, drinking patterns. One is considered regularly smoker if he or she has been smoking daily in the past 30 days. We differentiate marijuana from other hard drugs such as cocaine and crystal meth. Self-report of use of these marijuana and hard drugs are used to construct two dichotomous indicators. We define binge drinking as having 5 drinks in a row more than once a month for the past year. We also study an array of sex-related risky behaviors, including teenage sex, promiscuity, and unprotected sex.Specifically, one's first intercourse is considered a risky behavior, i.e. teenage sex, if it happens before age 15 Promiscuity is determined by the number of sex partners during the past 12 months, while we rely on self-report of use of condom or any contraceptives for the most recent intercourse. Finally, measures of delinquent behavior include self-report of serious or group fighting, use of weapon in fighting or robbery, gang membership, drinking while drunk, as well as arrest record.

Data Preparation   We used the core sample of Add Health, which is national representative. We subtract the 5% data which contains the highest percentage of missing data to obtain our final sample. We further decompose the data by race (whites, blacks and hispanics) and by gender.

3   Methodology

We use a latent variable model to fit multiple outcome variables of the Add Health data. Human skills, that is, cognitive, noncognitive and health abilities, are modeled as latent variables called factors. All outcomes are assumed to be affected by abilities, but in different ways.

3.1 The Model Description

Traditional models of factor analysis operate by loading data onto a set of latent variables, called factors. Identification procedures usually use rotation approaches and restrictions that do not account for the specific meaning of the variables being used. For our model, we perform a slightly different identification procedure. We use the meaning of individual variables to restrict the link between factors and data. Specifically, we construct factors with a structural interpretation by loading each factor onto a set of measures that attempt to quantify a distinct human quality.

We develop a model with cognitive, noncognitive, and health skills (factors). These skills are measured, with error, by variables such as cognitive tests, personality trait scales and health status. We use a distinct set of measures to identify each factor, and allow all factors to influence data outcomes. The parameters that govern the relation between factors and data outcomes are termed "factor loadings".

Notationally, let $\mathcal{J}$ be a an indexing set for the skills, so that $j \in \mathcal{J}$ represents skill j. Also, let $\boldsymbol{\Theta}_i$ be the vector of these skills for individual i. Let $\mathcal{K} \equiv \{1, \cdots, K\}$ be the set that indexes of all K outcomes, $Y_{k,i}$; $k \in \mathcal{K}$ (e.g. monthly earnings) for individual i. We represent the pooled vector of outcomes $Y_{k,i}$ over persons i by Y_k. The relation between an outcome $Y_{k,i}$ and a factor Θ_i^j is governed by α_k^j, which is the factor loading of ability j and outcome k.

Generally we can write our model as follows as

$$Y_{k,i} = f_k(\beta_k \cdot X_i + \alpha_k \cdot \boldsymbol{\Theta}_i, \epsilon_{k,i}),$$

where the function f depends on the outcome type. Again, $Y_{k,i}$ is the outcome k and $\boldsymbol{\Theta}_i$ is the vector of latent factors, both for person i. The vector of factor loadings associated with outcome k is represented by α_k. X_i are covariates,, such as gender and race, for person i. The error term $\epsilon_{k,i}$ is an i.i.d. random variable with mean zero.

For example, let the set of possible skills $\mathcal{J}$ contain the cognitive, noncognitive and health skills, that is $\mathcal{J} = \{c, n, h\}$ respectively. Let outcome Y_k be continuous and linear on the factors, thus, for person i,

$$Y_{k,i} = \beta_k \cdot X_i + \alpha_k^c \cdot \Theta_i^c + \alpha_k^n \cdot \Theta_i^n + \alpha_k^h \cdot \Theta_i^h + \epsilon_{k,i}.$$

Variable Specification We divide data into three categories: measures, outcomes, and covariates. The conceptual difference between the three may be seen in their relationship to the factors. Covariates are variables we wish to control for but do not direct influentiate the skills. Measures are variables that are specifically tied to one factor, for example, the Peabody Vocabulary Test (PPVT) is directly tied to the cognitive factor. Indded we say that the cognitive factor is measured with error by variables such as PPVT. Finally, outcomes are variables that may are influentiated by all factors. examples of outcomes are later life drug use, schooling decisions, employment status or risk behavior variables. In our models, these differences are represented by differences in restrictions on the factor loadings, which are the parameters that tie the factors and variables as mentioned. Outcomes are loaded onto all factors, while each measure is loaded only onto the factor to which it particularly relates. In other words, for measures, all factor loadings are set to zero except for one. Notationally, let $M_{k,i}^{j}$ denote the k-th measure for skill j and person i, so if the relation between the measure and teh factor is linear, that is:

$$M_{k,i}^{j} = \beta_k \cdot X_i + \alpha_k^c \cdot \Theta_i^c + \alpha_k^n \cdot \Theta_i^n + \alpha_k^h \cdot \Theta_i^h + \epsilon_{k,i}.,$$

then, $\alpha_k^{j'} = 0 \ \forall j' \neq j \ ; \ j, j' \in \mathcal{J}$.

Measures are always modeled as linear on the factors. For all factors $j \in \mathcal{J}$, define a set of size M_j, $\mathcal{M}_j \equiv \{1, ..., M_j\}$ to be the set of all measures for the factor j. We restrict the factor loading for the first measure, M_1^j, to be 1. Thus, for we may write the measures as:

$$M_{1,i}^{j} = X_i \cdot \beta_1^j + 1 \cdot \Theta_i^j + \epsilon_{1,i}^j$$

$$M_{m,i}^{j} = X_i \cdot \beta_m^j + \varphi_m^j \cdot \Theta_i^j + \epsilon_{m,i}^j, \ m \in \mathcal{M}_j \setminus \{1\}, \ \forall j \in \mathcal{J}.$$

Outcomes, however, may be continuous, categorical, or binary variables, so we model them in a number of ways. We use linear models for continuous variables, multinomial probits for categorical variables, and probits for binary variables. Thus, for continuous variables, we have

$$Y_k = X \cdot \beta_k + \sum_{j \in \mathcal{J}} \alpha_k^j \cdot \boldsymbol{\Theta}^j + \epsilon_k$$

Alternatively, for a categorical variables, where the number of possible values (usually the number of choices) of the categorical variable C_k is $\mathcal{C}$. The probability of choosing the choice $c \in \{1, \cdots, \mathcal{C}\}$ is given by a multinomial probit:

$$Pr(C_{k,i} = c) = Pr(max(\omega_{c',i} : c' = 1, \cdots, C) = c)$$

$$[\omega_{1,i}, \cdots, \omega_{C,i}] \sim N(\mu_i, \Sigma); \mu_i = [\mu_{1,i}, \ldots, \mu_{C,i}]'$$

$$\mu_{c,i} = X_i \cdot \beta_{k,c} + \alpha_{k,c} \cdot \boldsymbol{\Theta}_i \; \forall c \in \{1, \cdots, \mathcal{C}\}$$

Identification of the model are discussed in the appendix.

For a binary variable, we use a probit model, thus

$$Prob(Y_{k,i} = 1) = Pr\left(X_i \cdot \beta_k + \sum_{j \in \mathcal{J}} \alpha_k^j \cdot \boldsymbol{\Theta}_i^j > \epsilon\right); \; \epsilon \sim N(0,1)$$

Finally, we define our factors as follows: Let $\mathcal{J} = c, n, h$, denote cognitive, noncognitive, and health factors. Let the vector of factors ordered as c, n, h be represented by $\boldsymbol{\Theta} = []$. Our first assumption is that the $\boldsymbol{\Theta}$ is distributed as a mixture of normals with G components, that is:

$$\boldsymbol{\Theta} \sim \sum_{g=1}^{G} P_g \odot N(\mu_g, \boldsymbol{\Sigma}_g) : \sum_{g=1}^{G} P_g = 1 \text{ and } P_g > 0 \, \forall \; g \in \{1, \cdots, G\}.$$

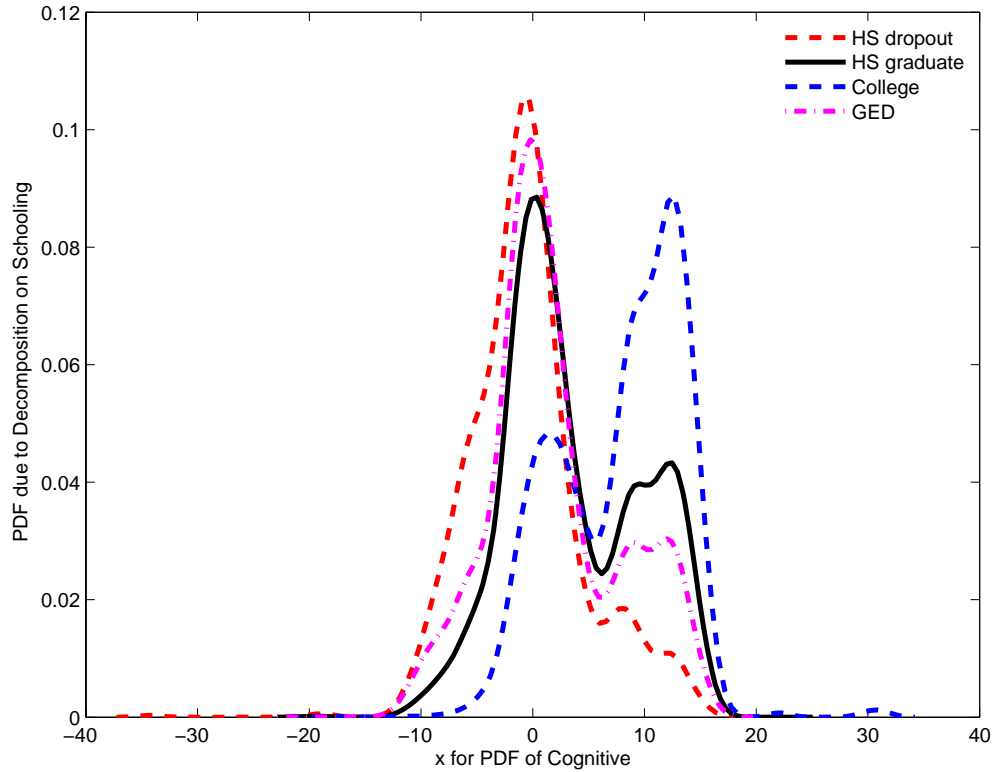The log-likelihood of the model and identification are described in Appendix W.1.

4 Empirical results

Data

Estimates Conditioned on Schooling Status   A special outcome is schooling status, which is divided into four categories: dropout, GED graduate, high school graduated and college graduated. We analyze the factor and outcomes distributions conditional on each of the schooling decision.

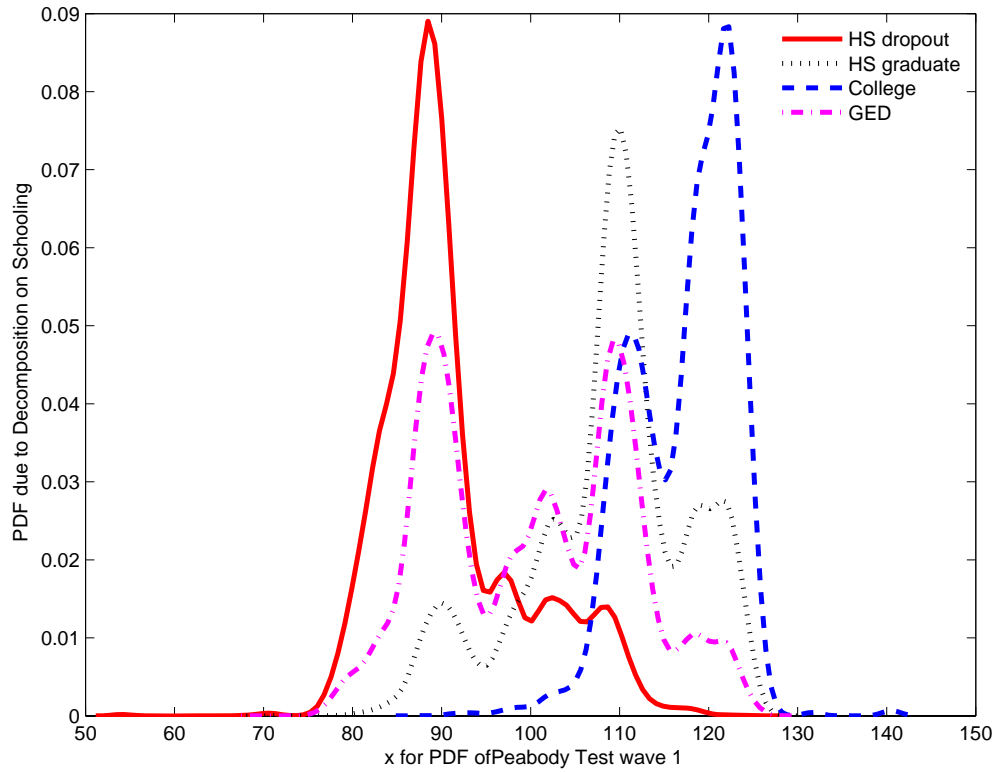**Figure 1:** Cognitive Ability as Latent Factor

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

7

Figure 2: Peabody Vocabulary Test Score, Wave I
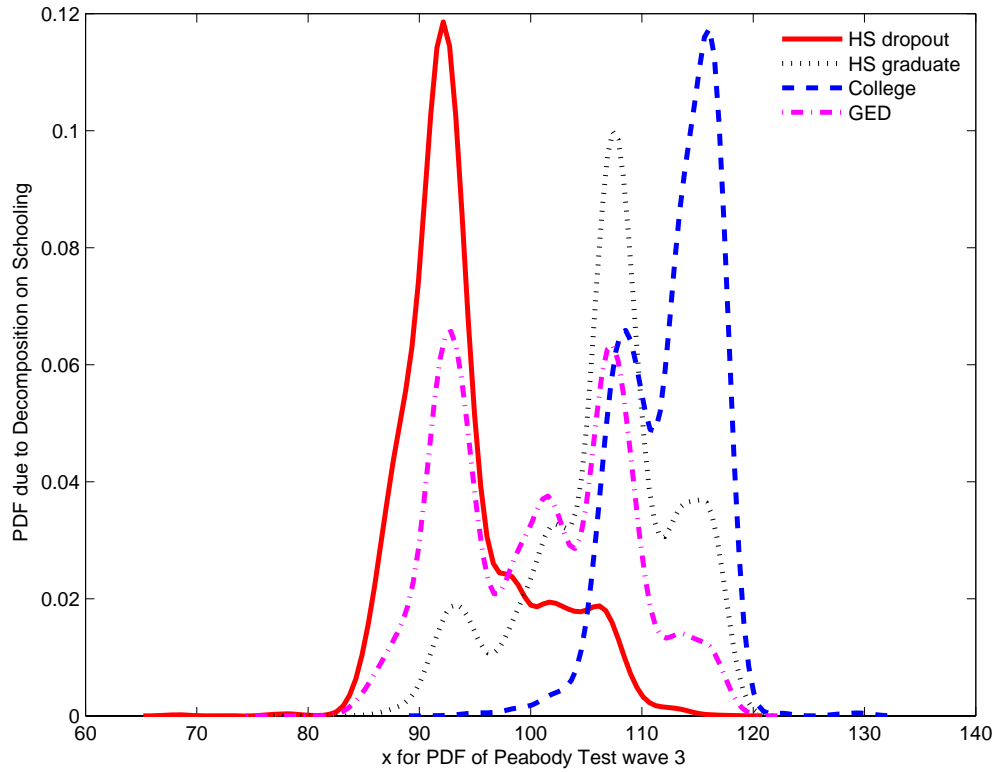
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 3: Peabody Vocabulary Test Score, Wave III
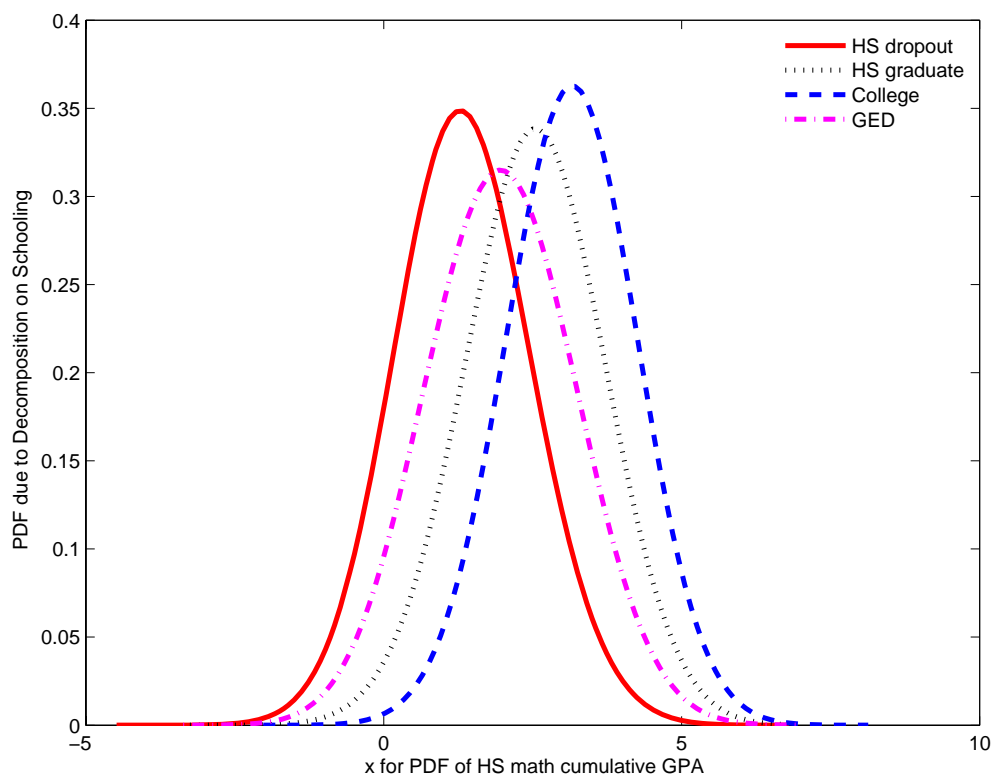
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

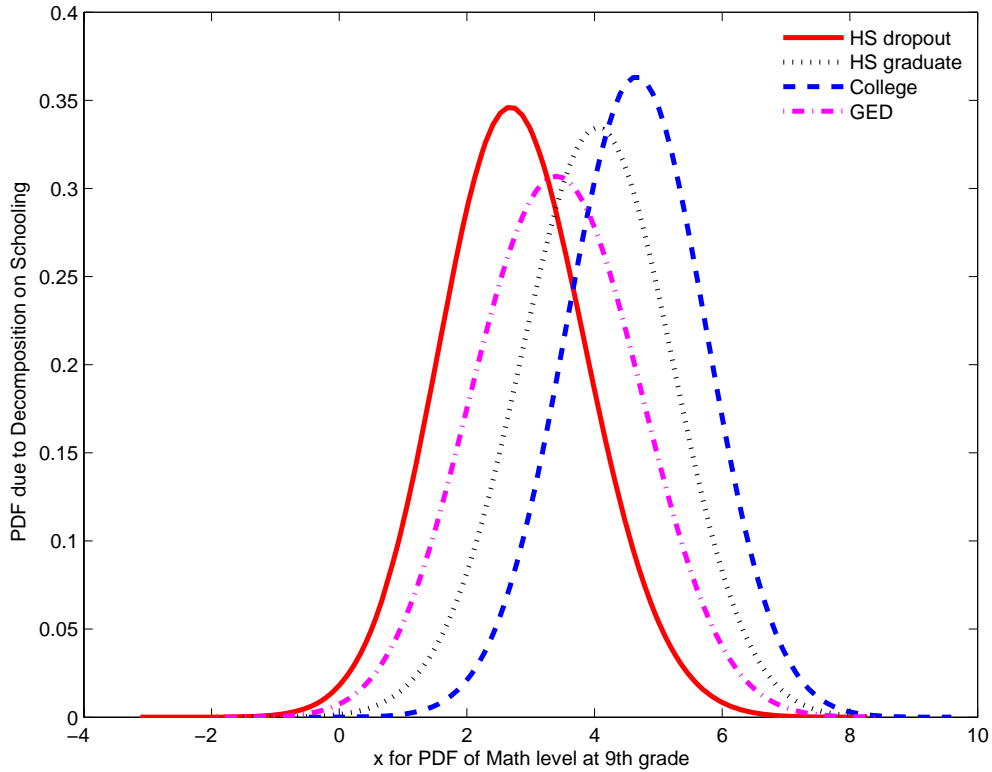**Figure 4:** High School Math GPA (Cumulative)

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

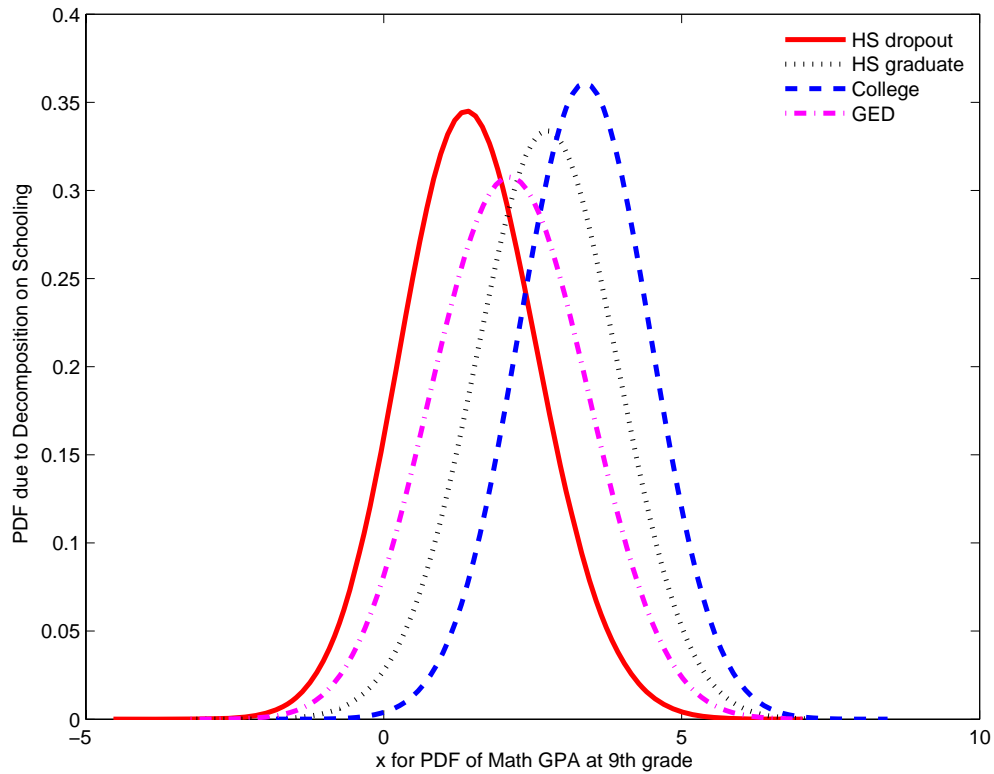**Figure 5:** Level of Math Courses Taken at 9th Grade

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

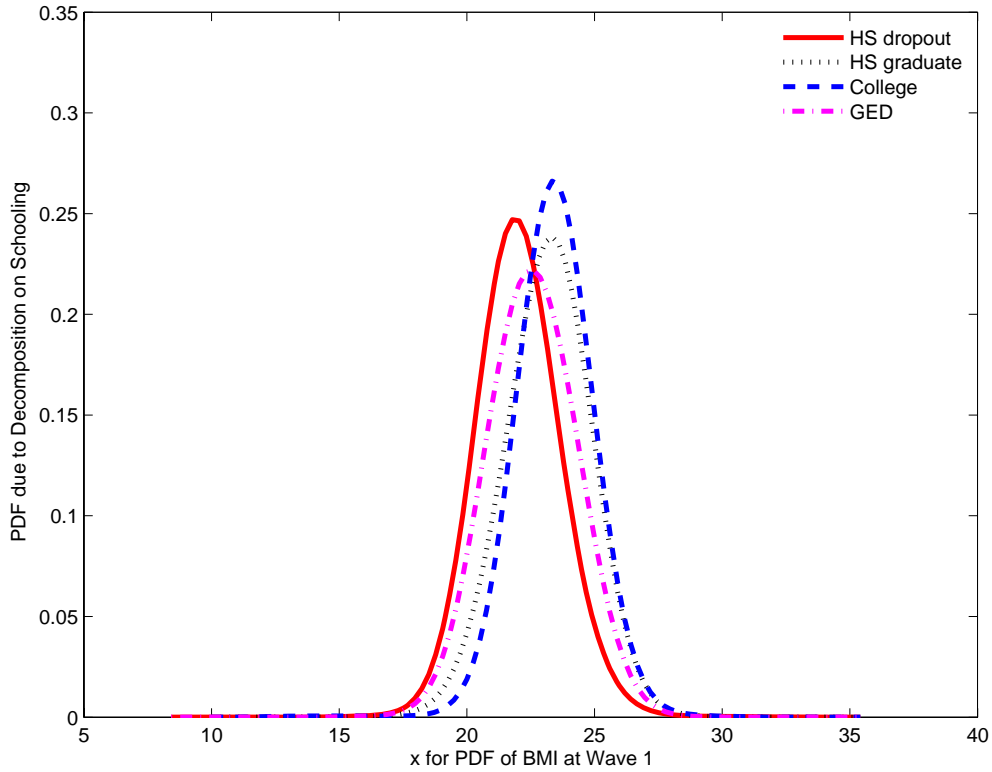**Figure 6:** Math GPA at 9th Grade , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

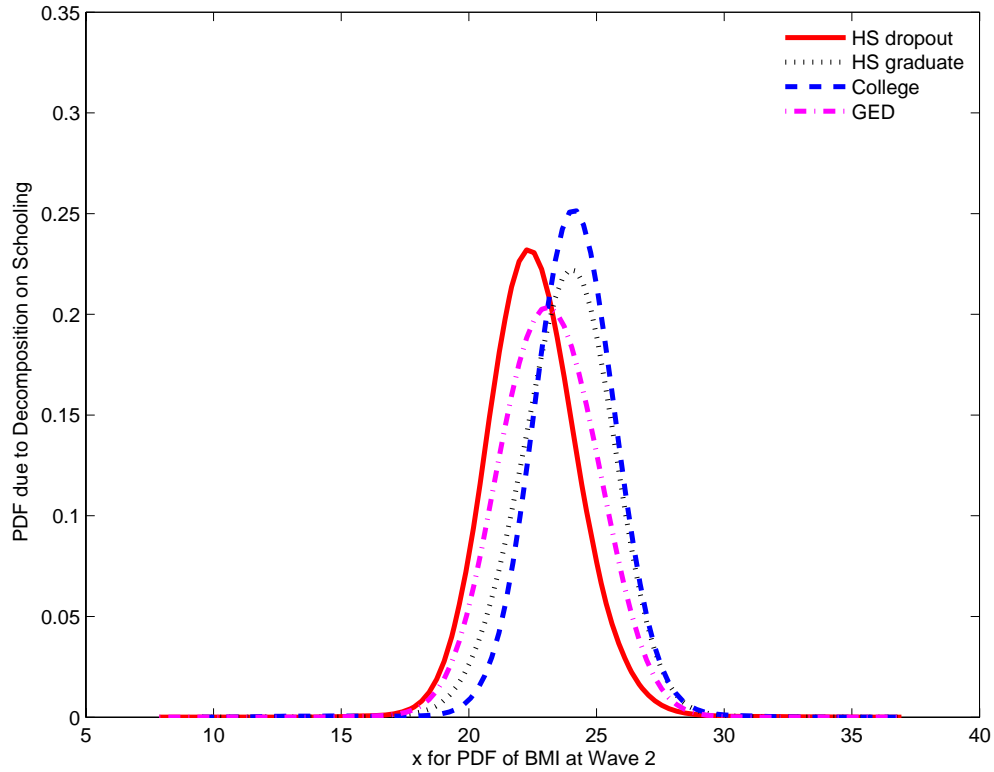**Figure 7:** Body Mass Index (BMI) , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

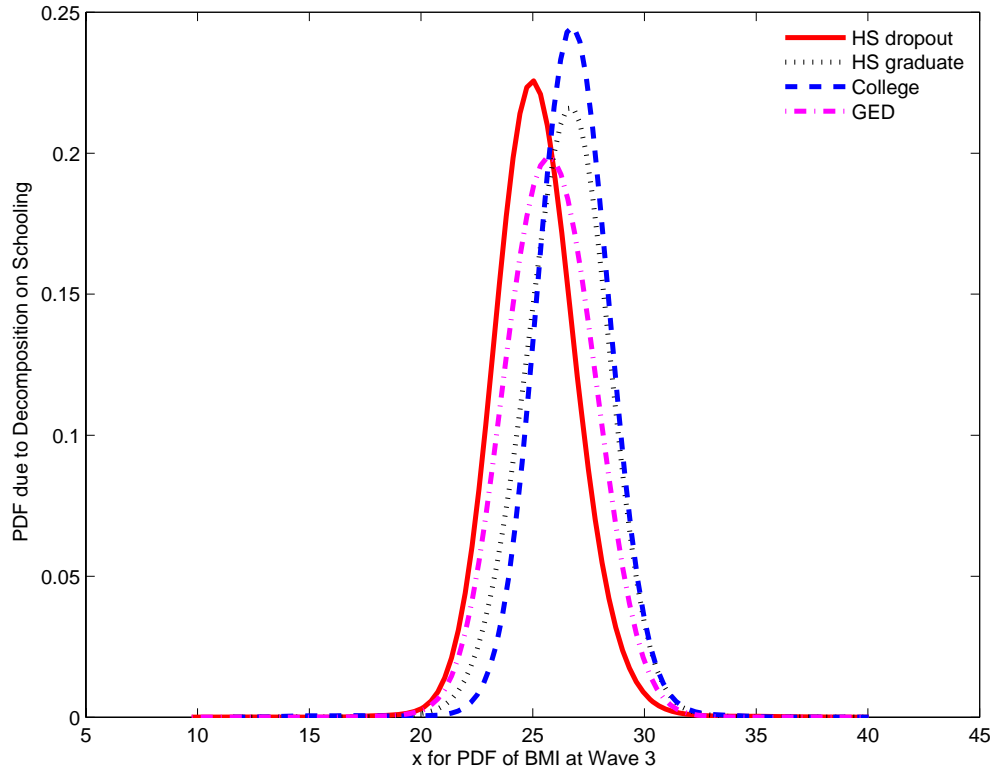**Figure 8:** Body Mass Index (BMI) , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

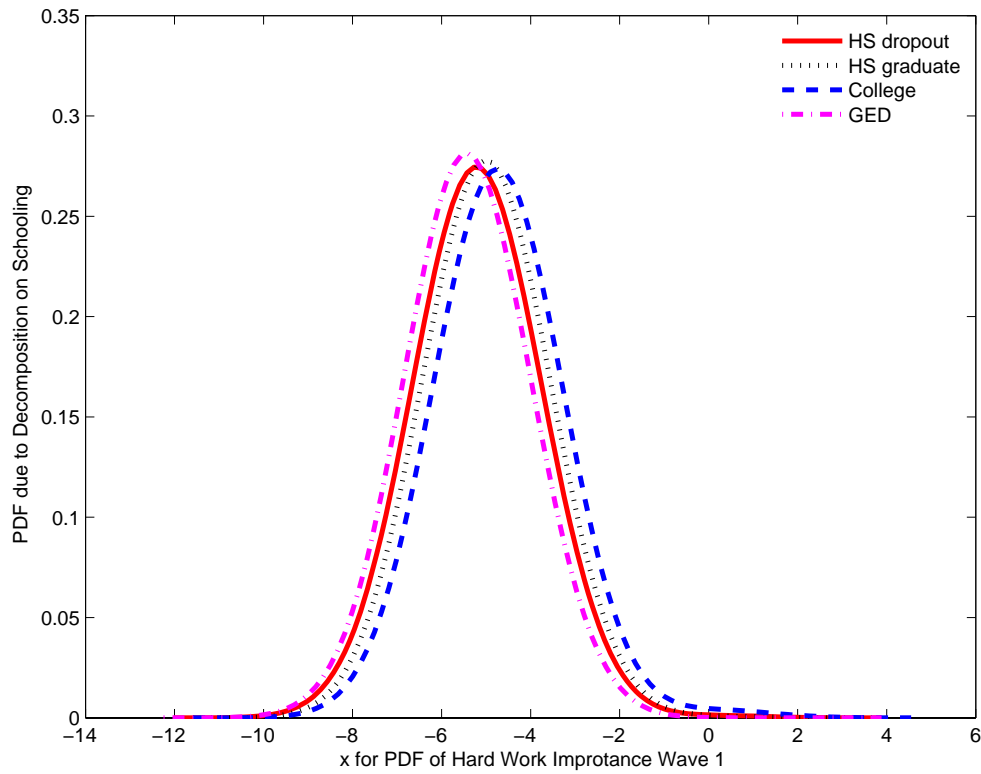**Figure 9:** Body Mass Index (BMI) , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

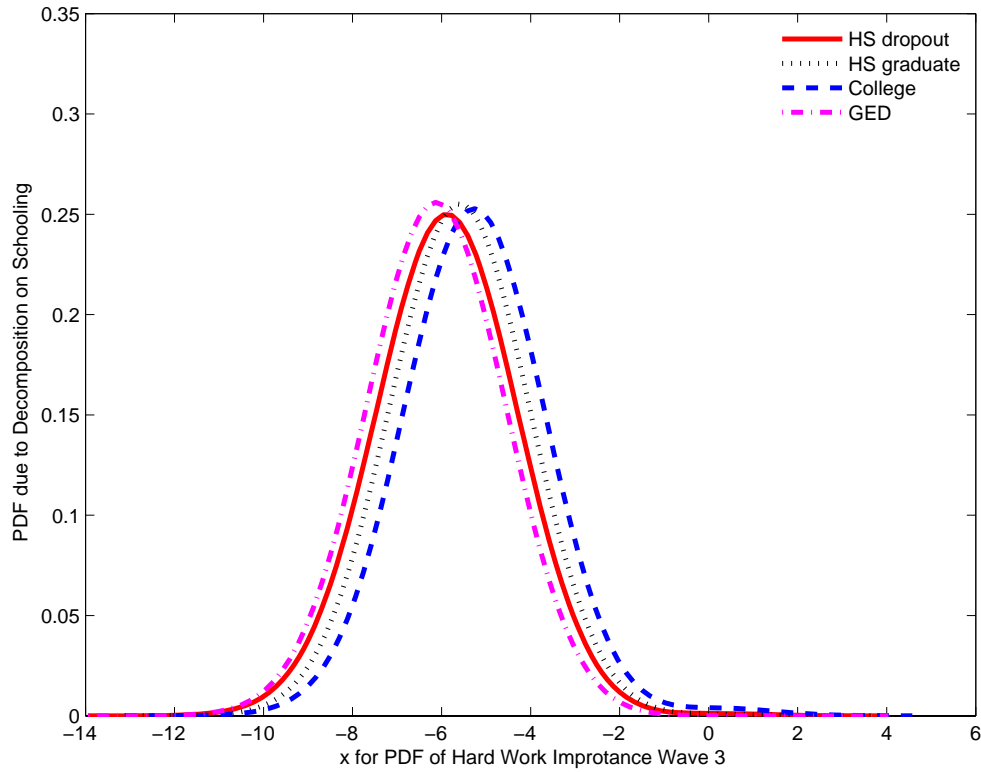**Figure 10:** Belief in Accomplishment through Hard Work , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

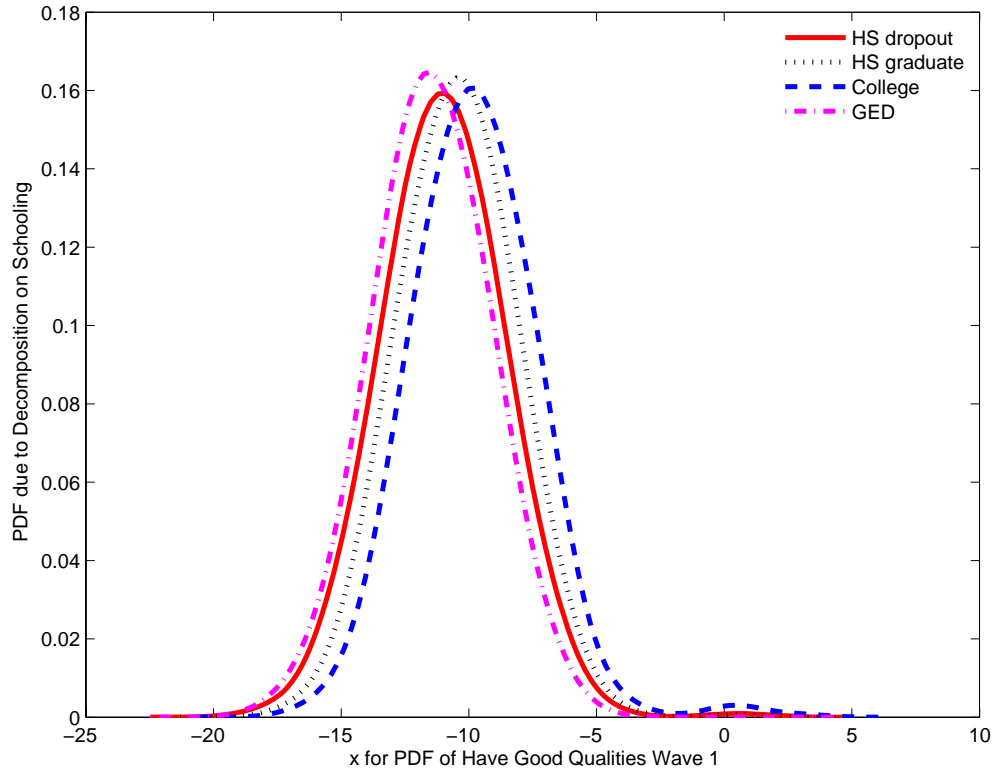**Figure 11:** Belief in Accomplishment through Hard Work , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

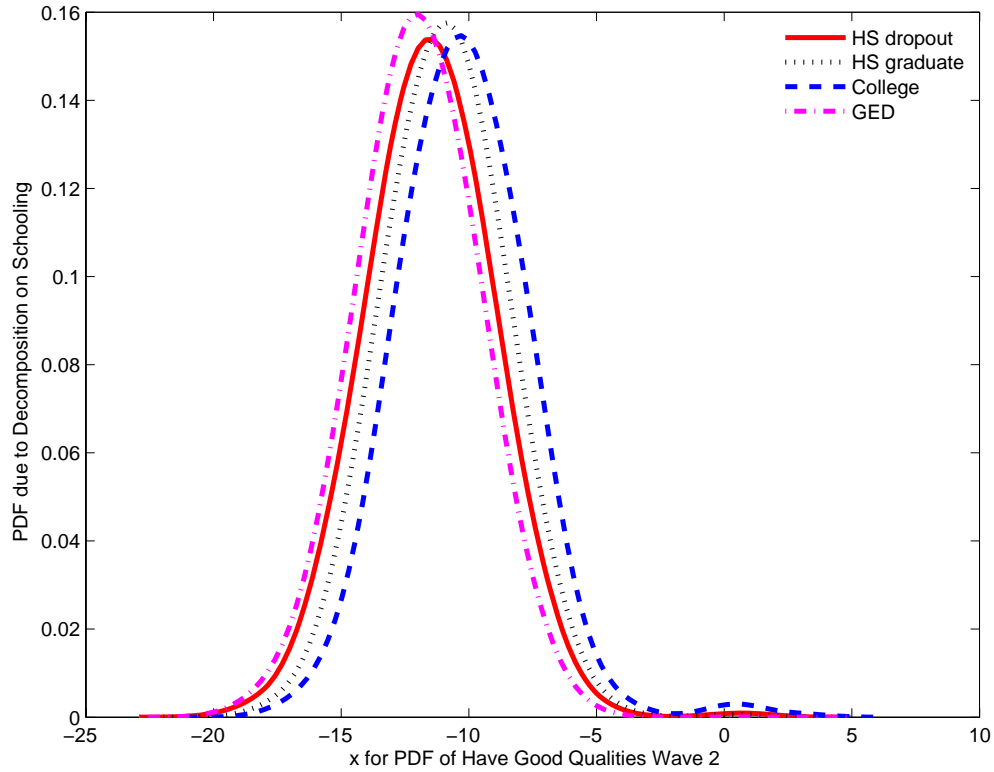**Figure 12:** "You Have A Lot of Good Qualities" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

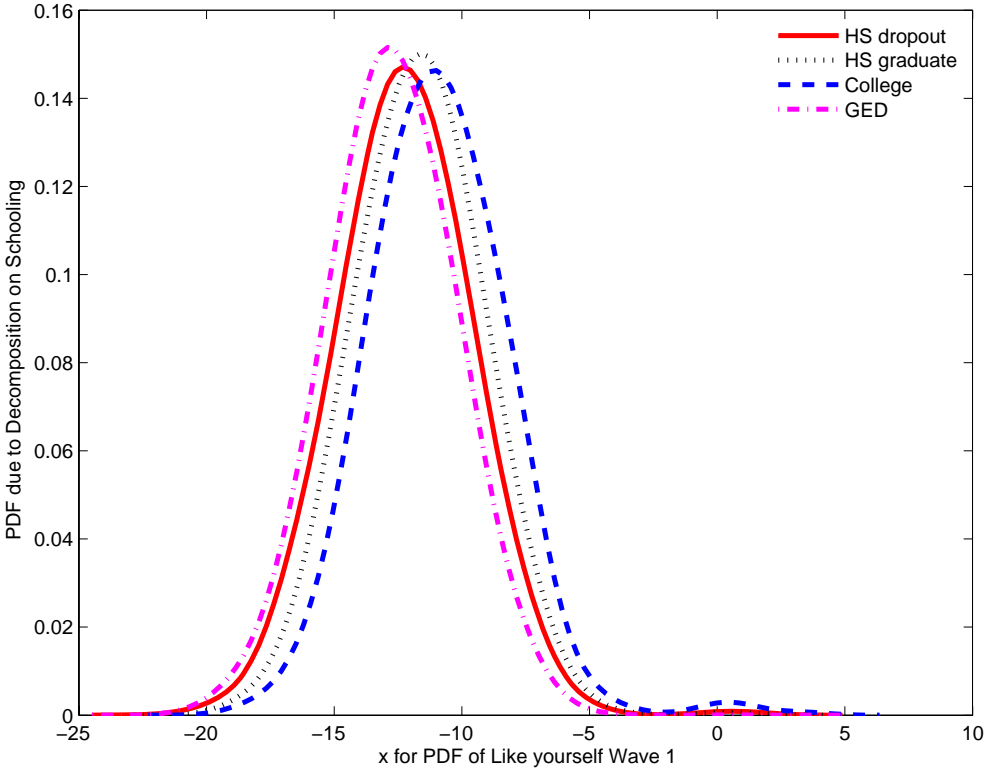**Figure 13:** "You Have A Lot of Good Qualities" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

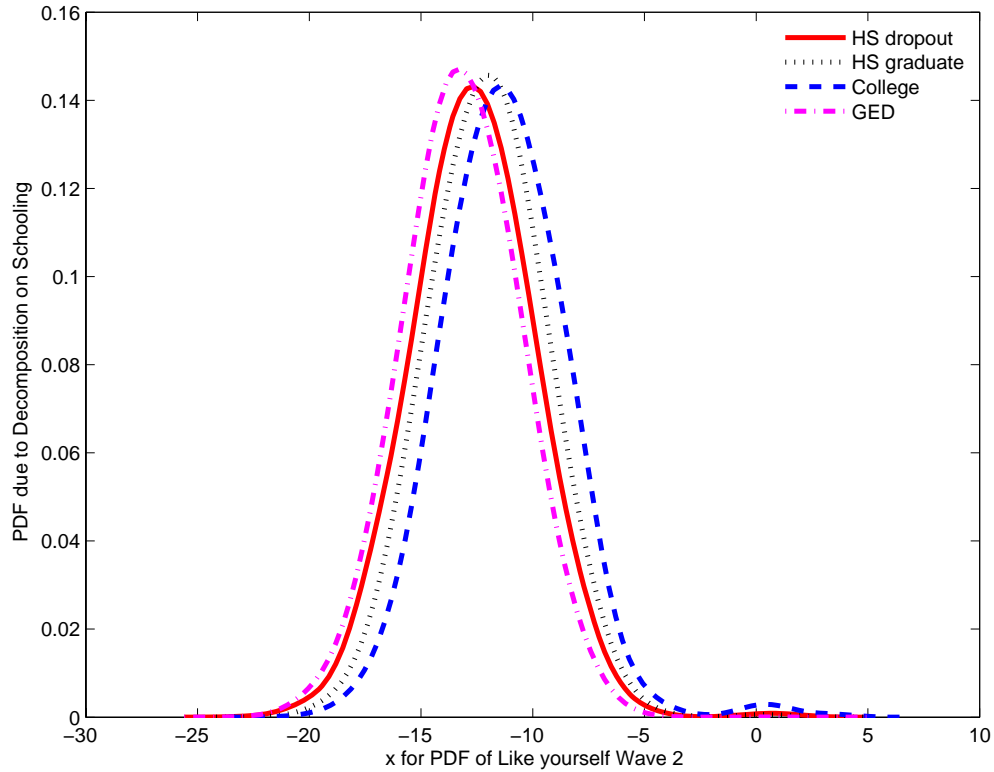**Figure 14:** "You Like Yourself" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

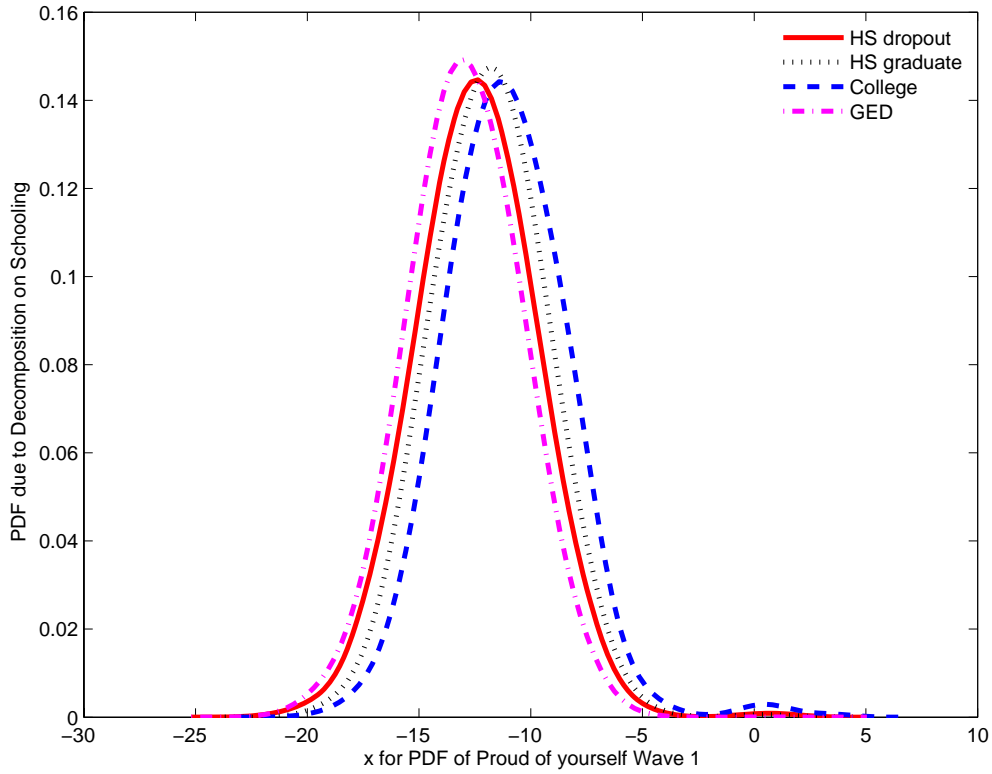**Figure 15:** "You Like Yourself" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

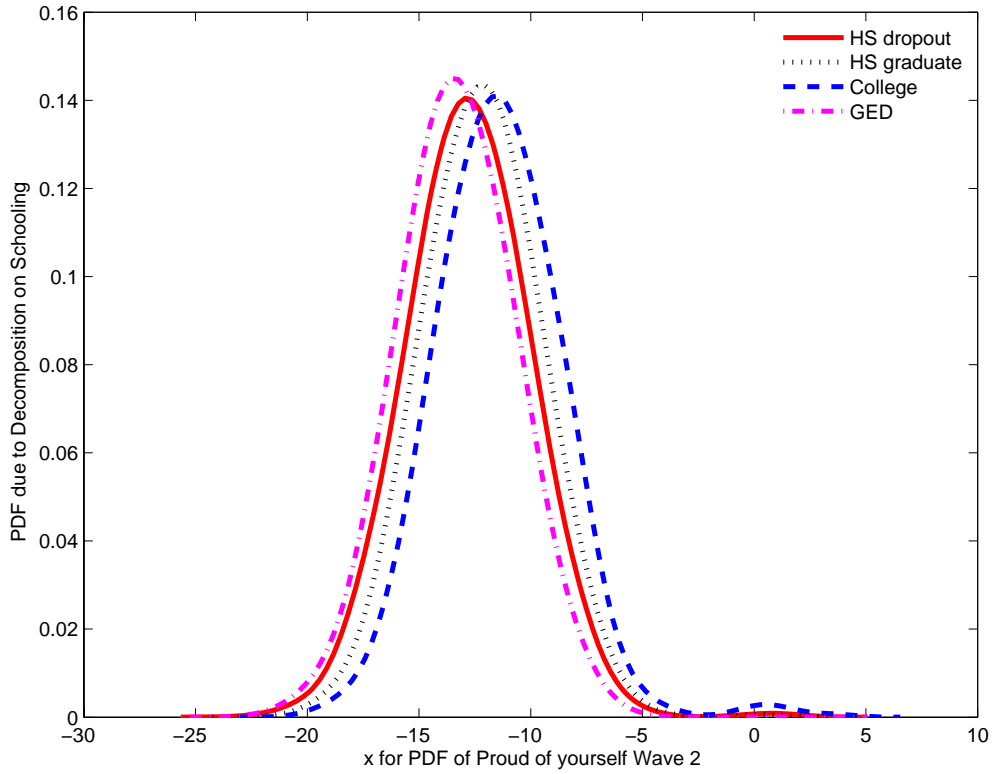**Figure 16:** "You Are Proud of Yourself" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

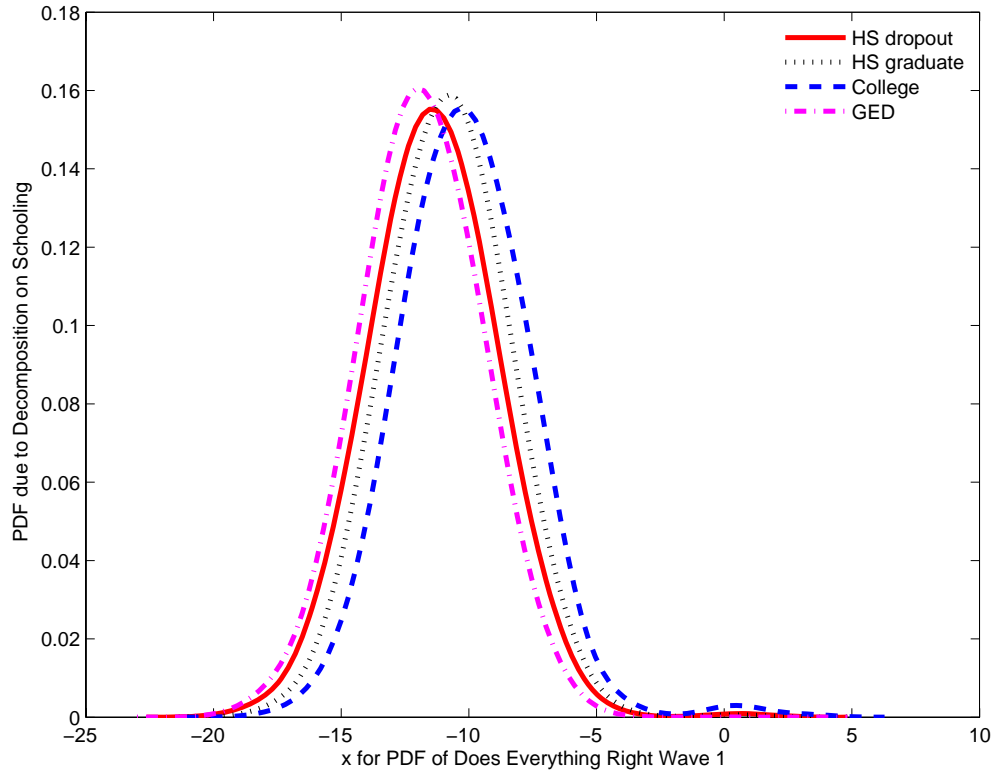**Figure 17:** "You Are Proud of Yourself" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

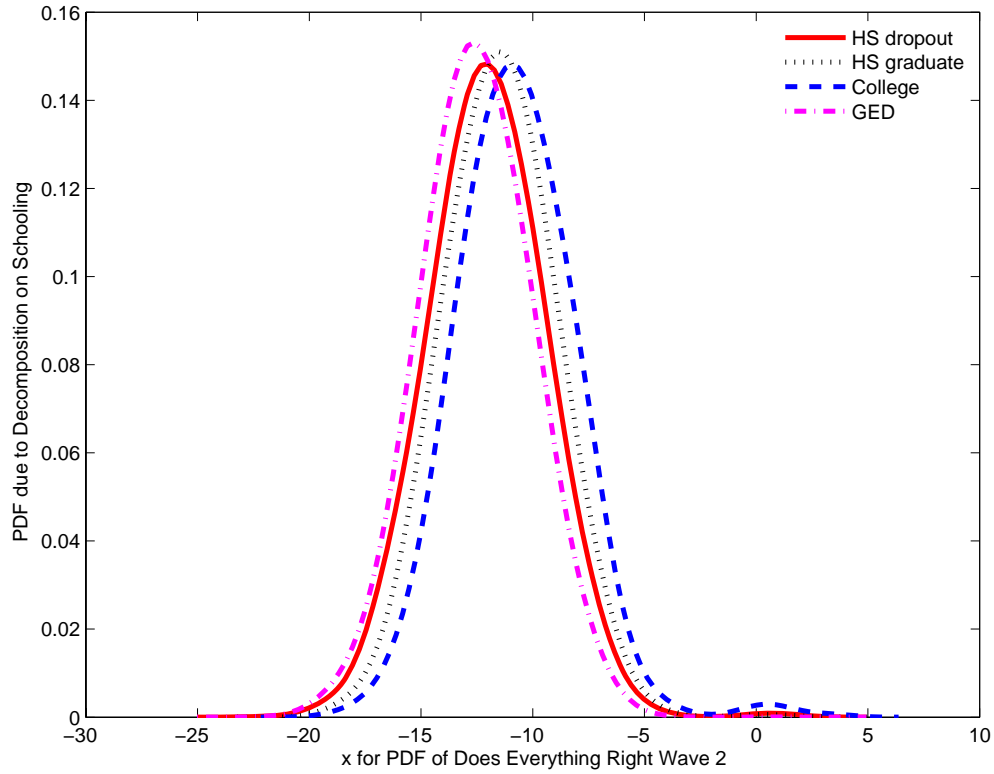**Figure 18:** " You feel You're Doing Everything Right" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

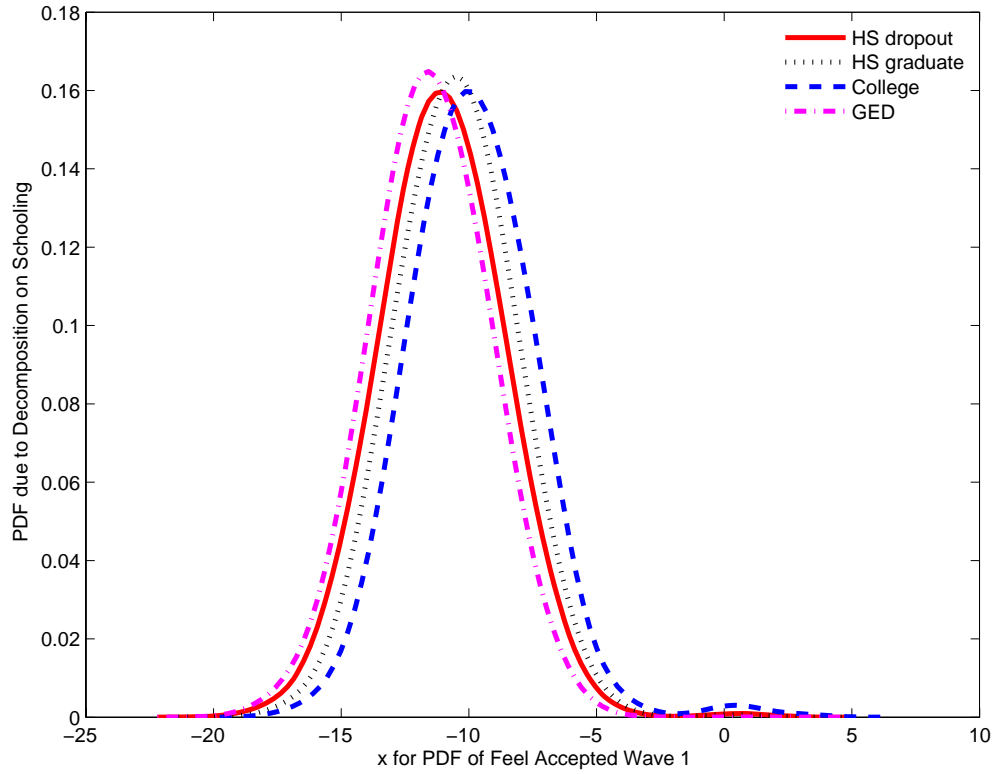**Figure 19:** " You Feel You're Doing Everything Right" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

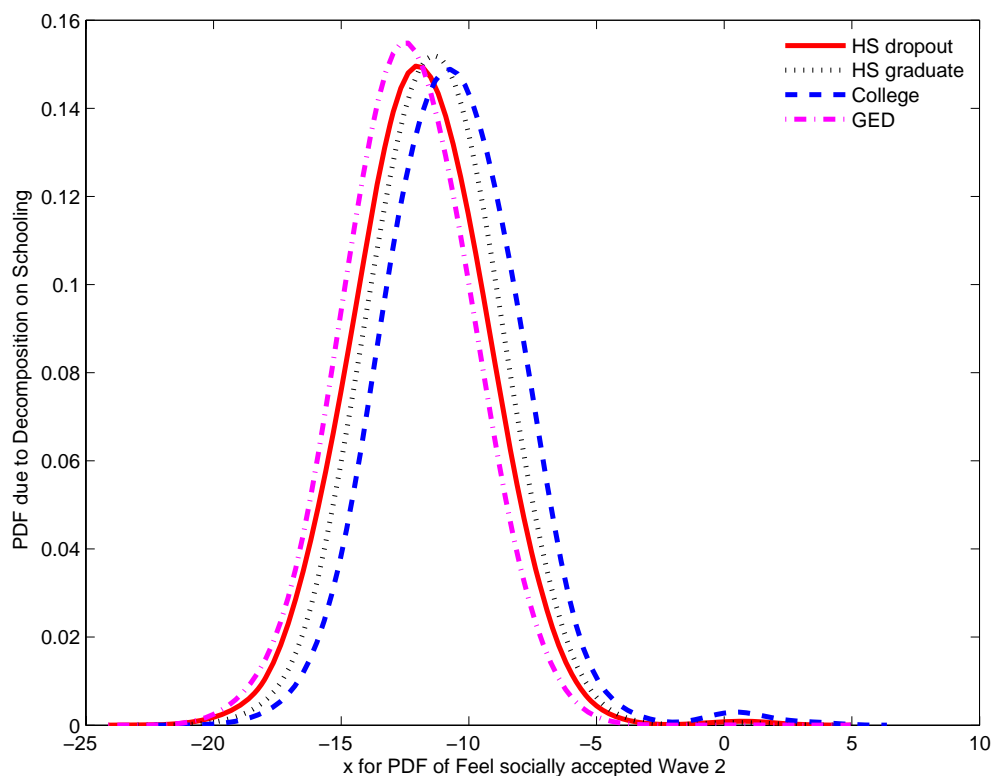**Figure 20:** "You Feel Socially Accepted" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

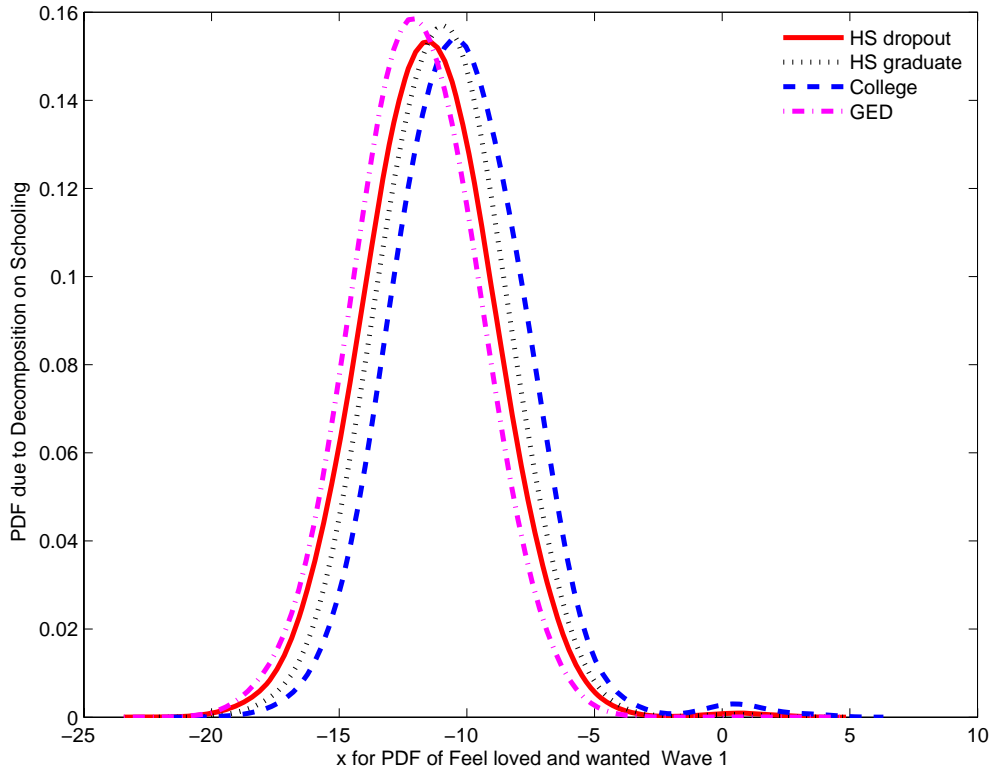**Figure 21:** "You Feel Socially Accepted" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

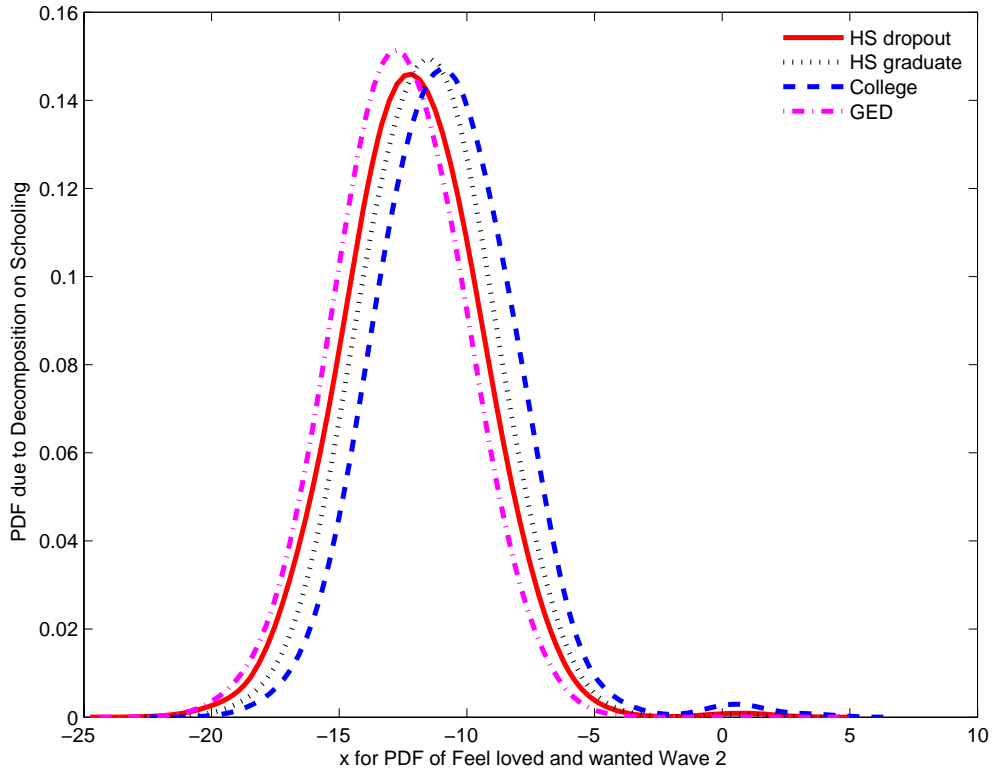**Figure 22:** "You Feel Loved and Wanted" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

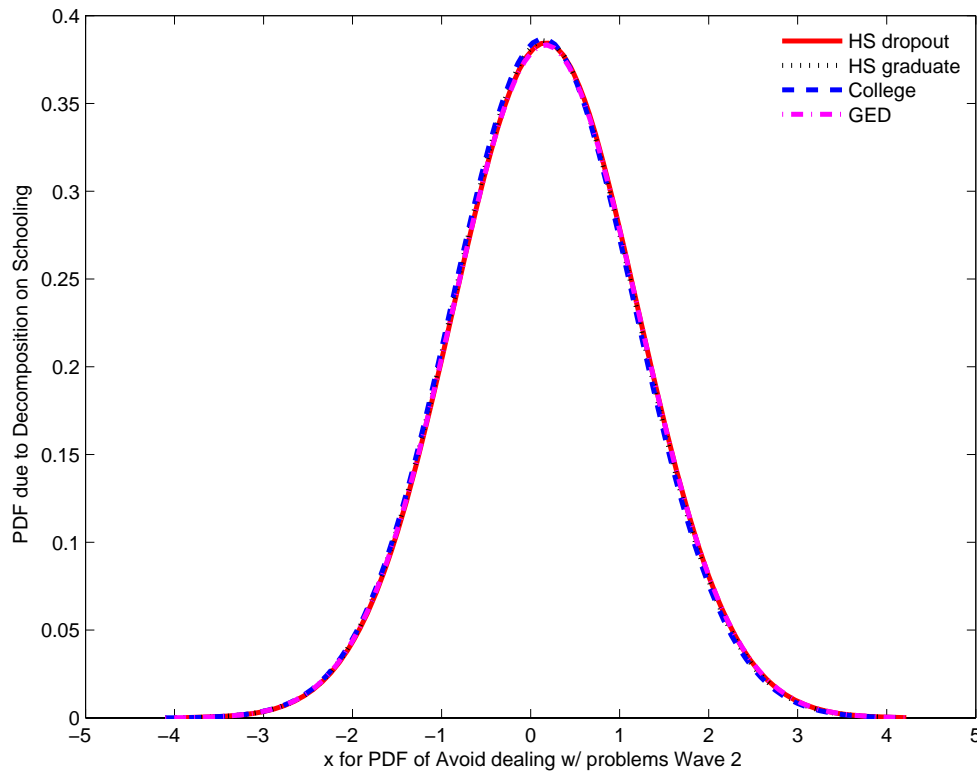**Figure 23:** "You Feel Loved and Wanted" , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 24: Always Avoid Dealing with Problems in Your Life , Wave II
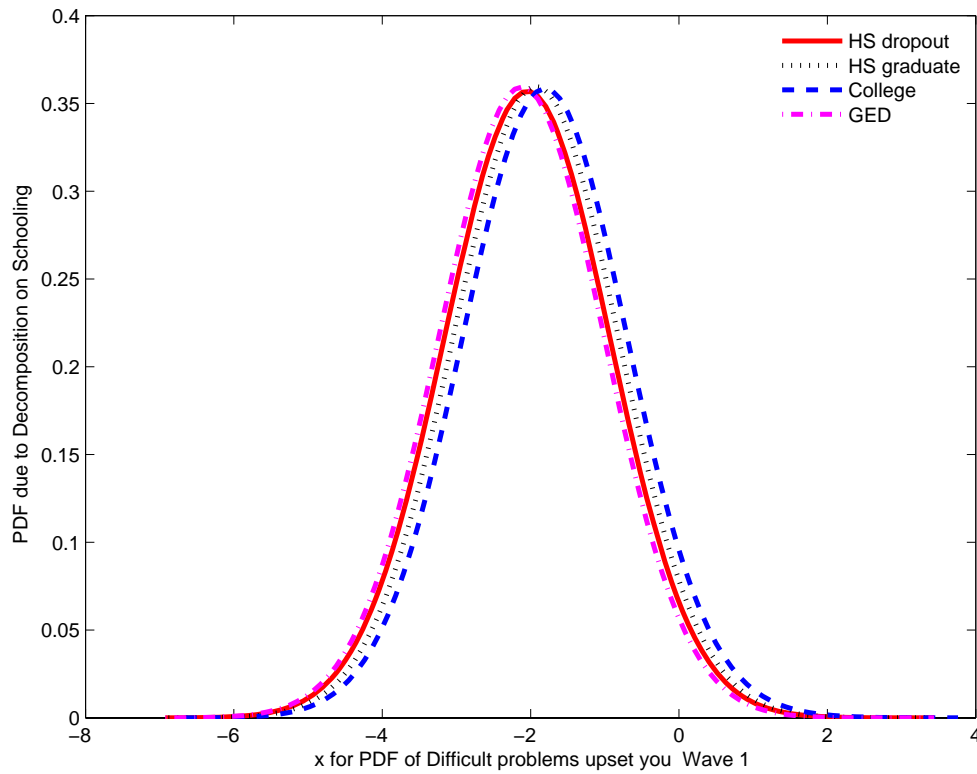
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

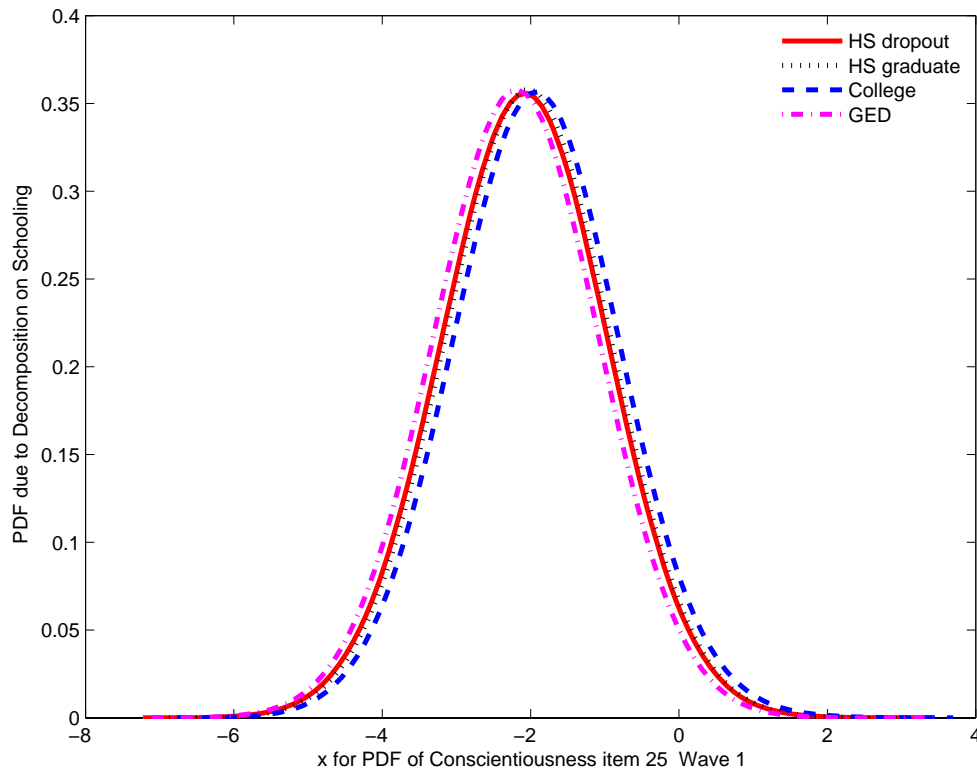**Figure 25:** Difficult Problems Make Your Upset , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
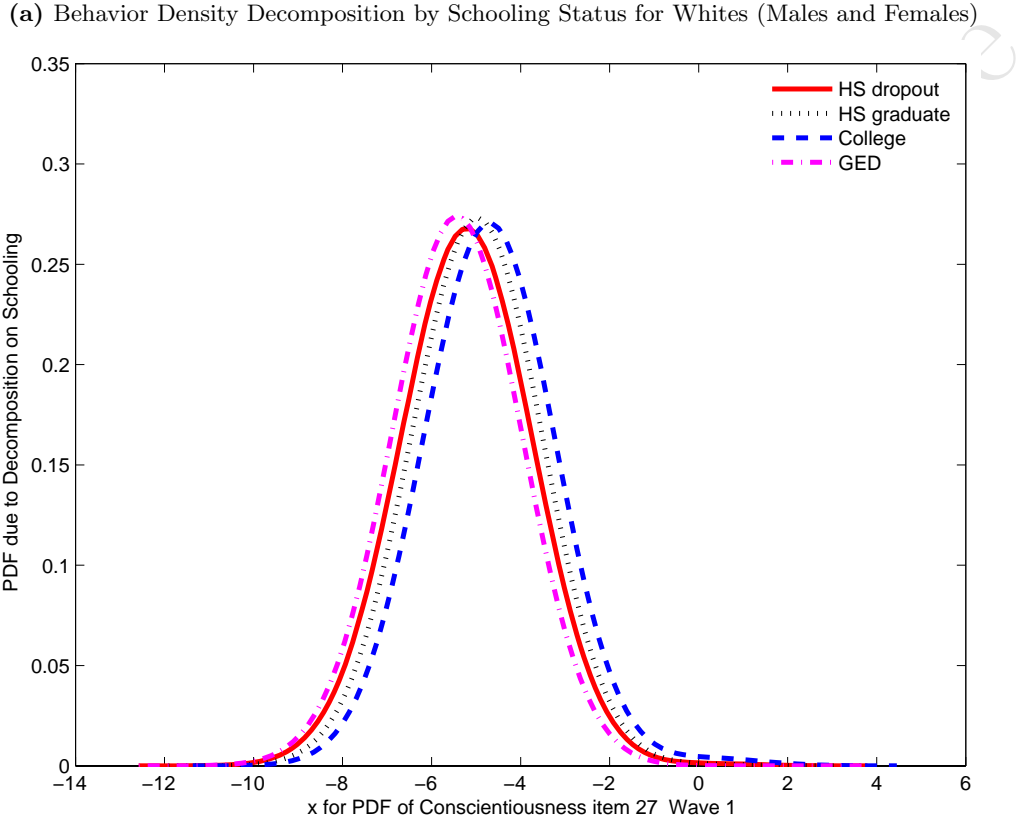
Figure 26: Conscientiousness: Make Decision based on "Gut Feeling" , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 27: Conscientiousness: You Try Different Ways to Solve Problems , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

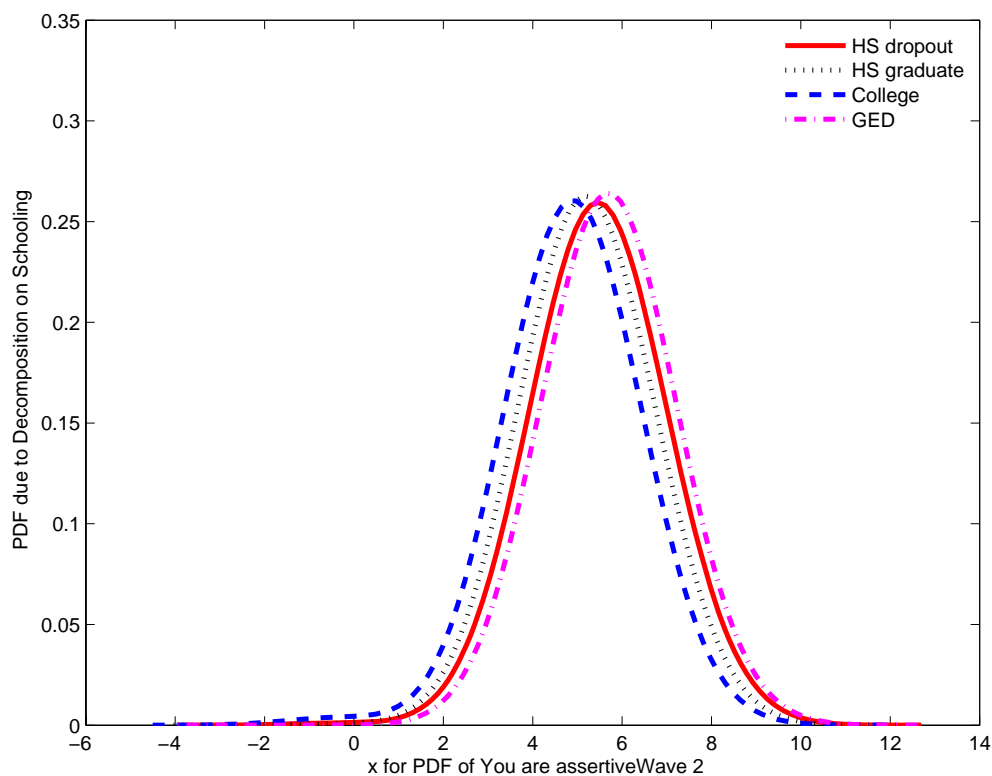**Figure 28:** Never Criticize People , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
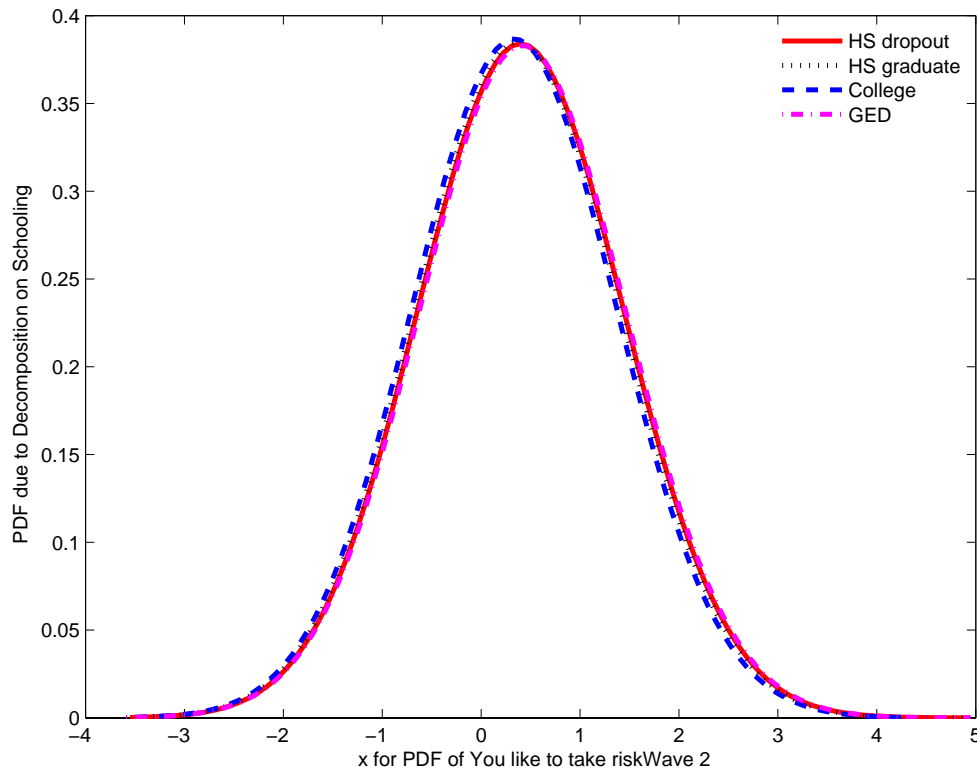
34

Figure 29: Assertive , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

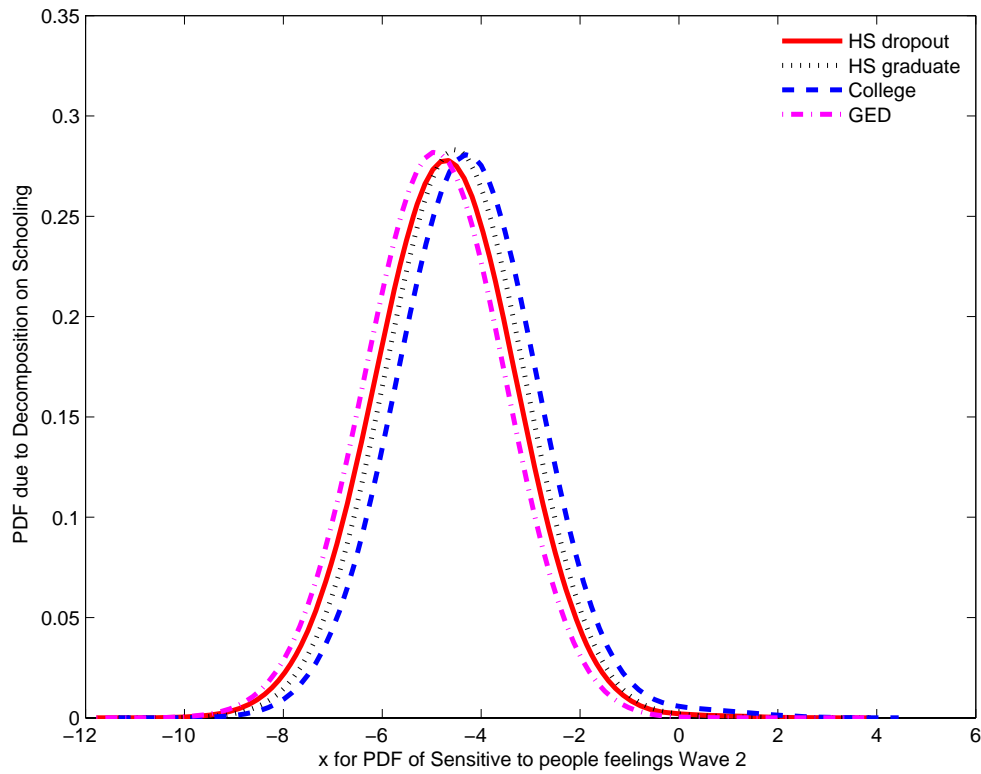**Figure 30:** Like to Take Risk , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

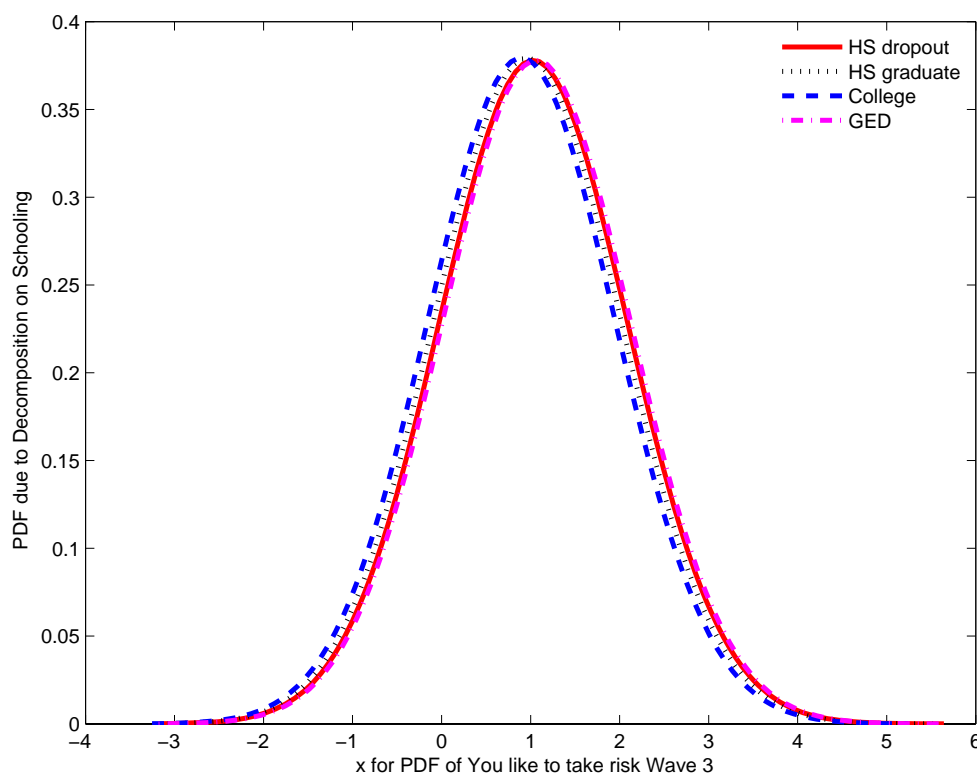**Figure 31:** Sensitive to People's Feelings , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

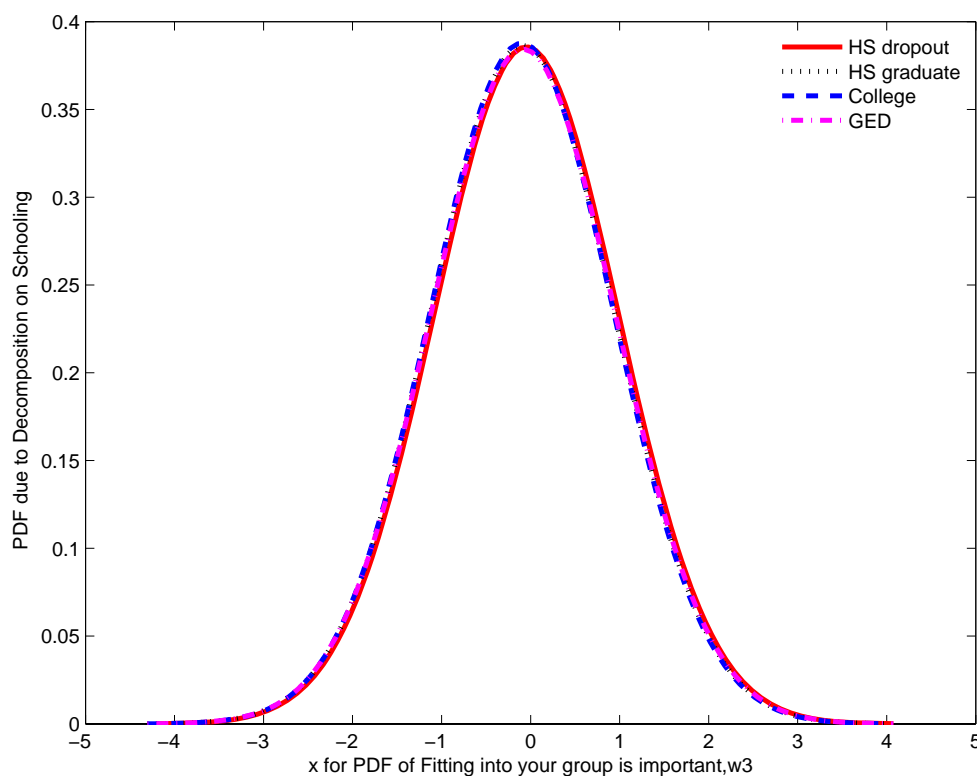**Figure 32:** Like to Take Risk , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

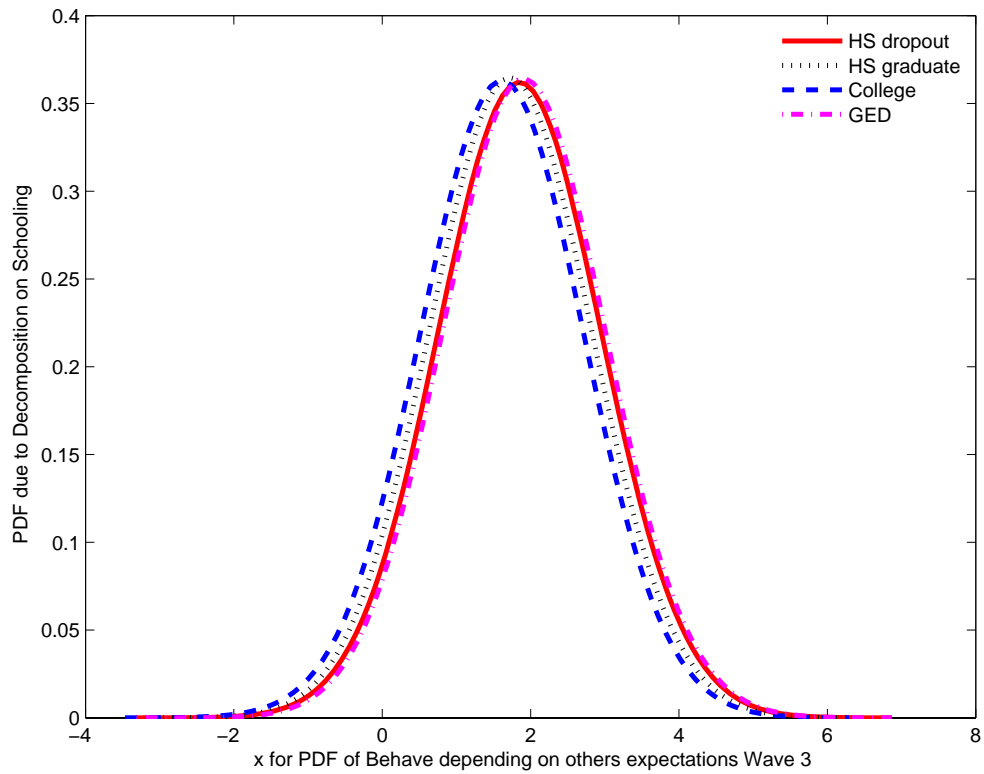**Figure 33:** "Fitting in Your Group is Important" , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 34: Behave Depending on Other People's Expectations , Wave III
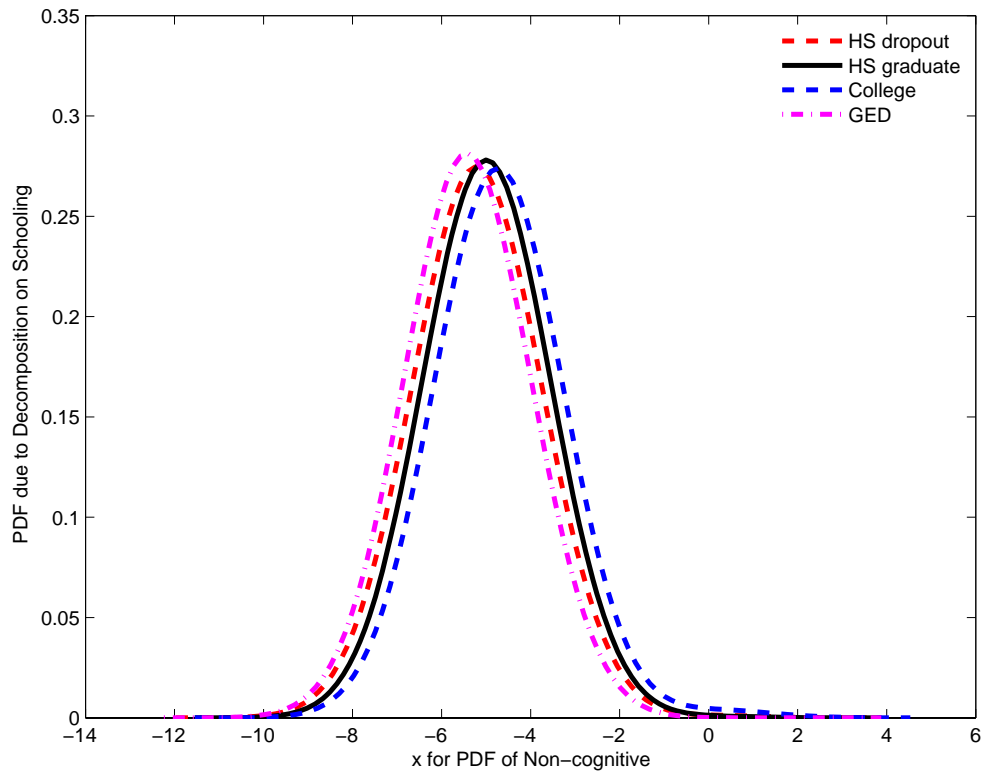
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

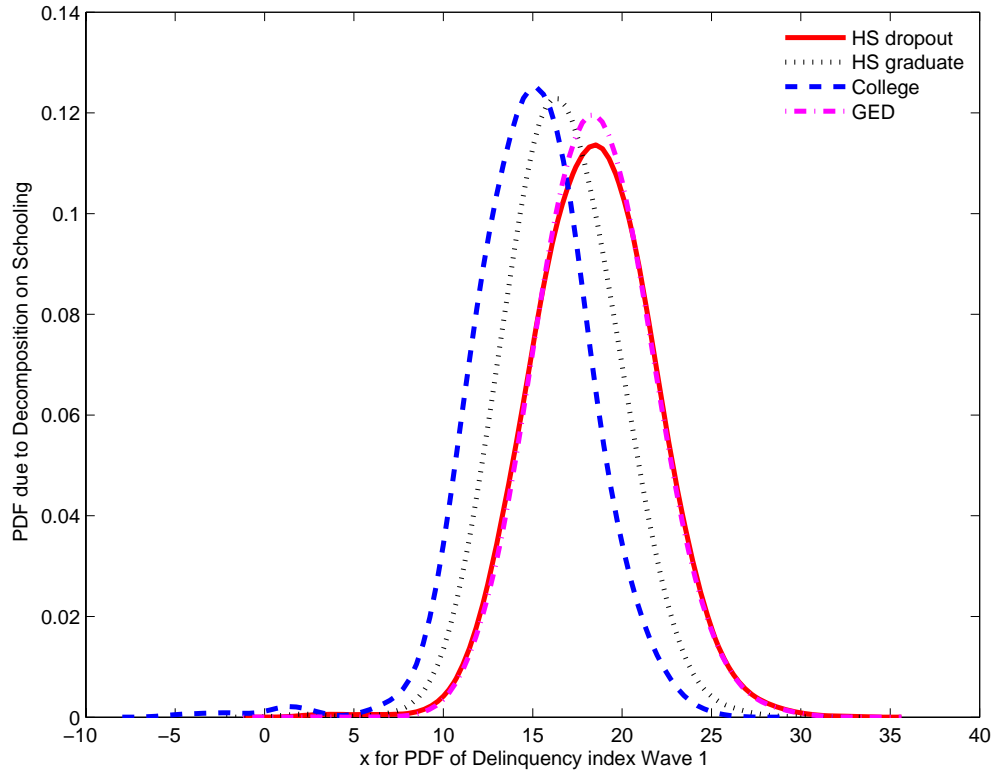**Figure 35:** Noncognitive Skill as Latent Factor, Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 36: Delinquency Index , Wave I
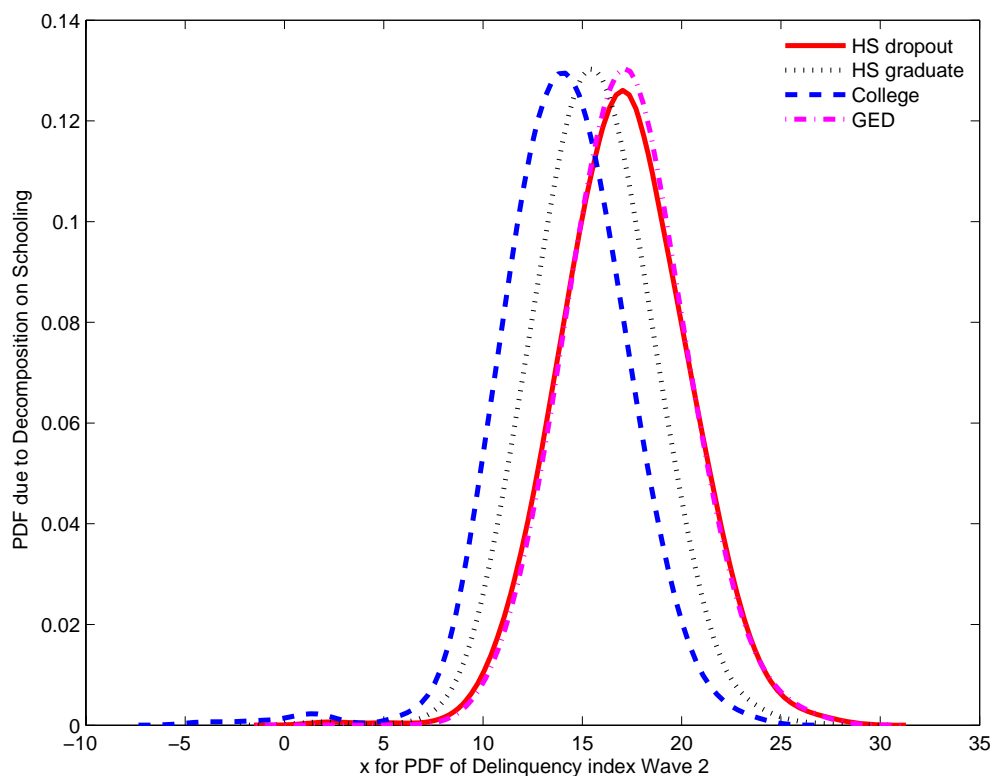
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

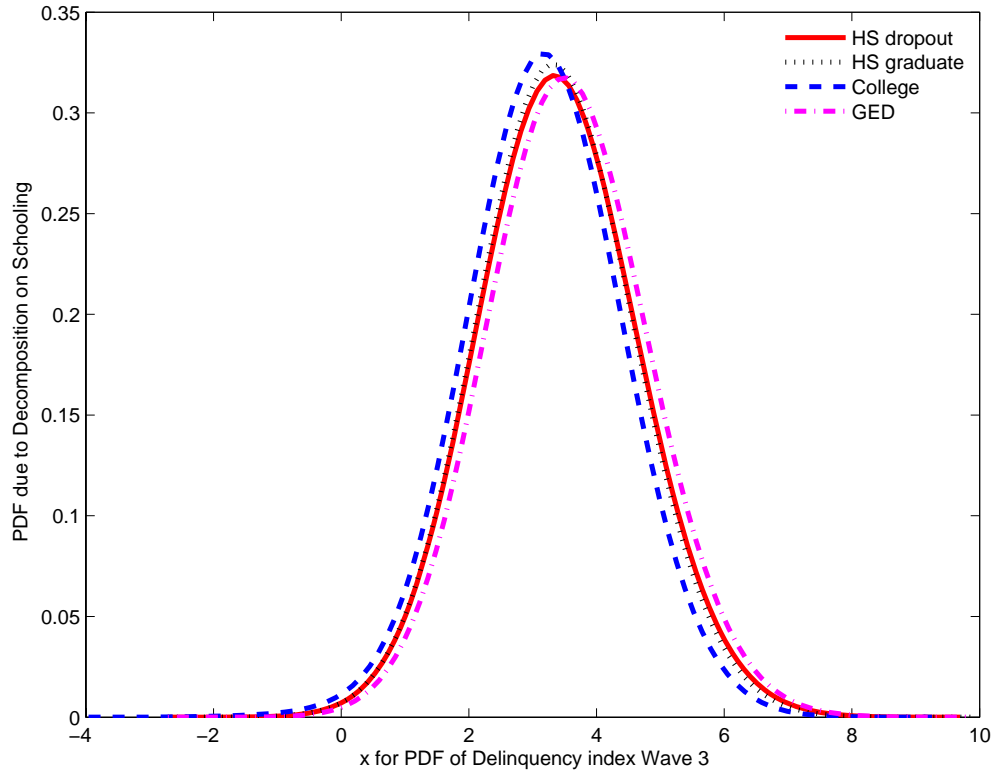**Figure 37:** Delinquency Index , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

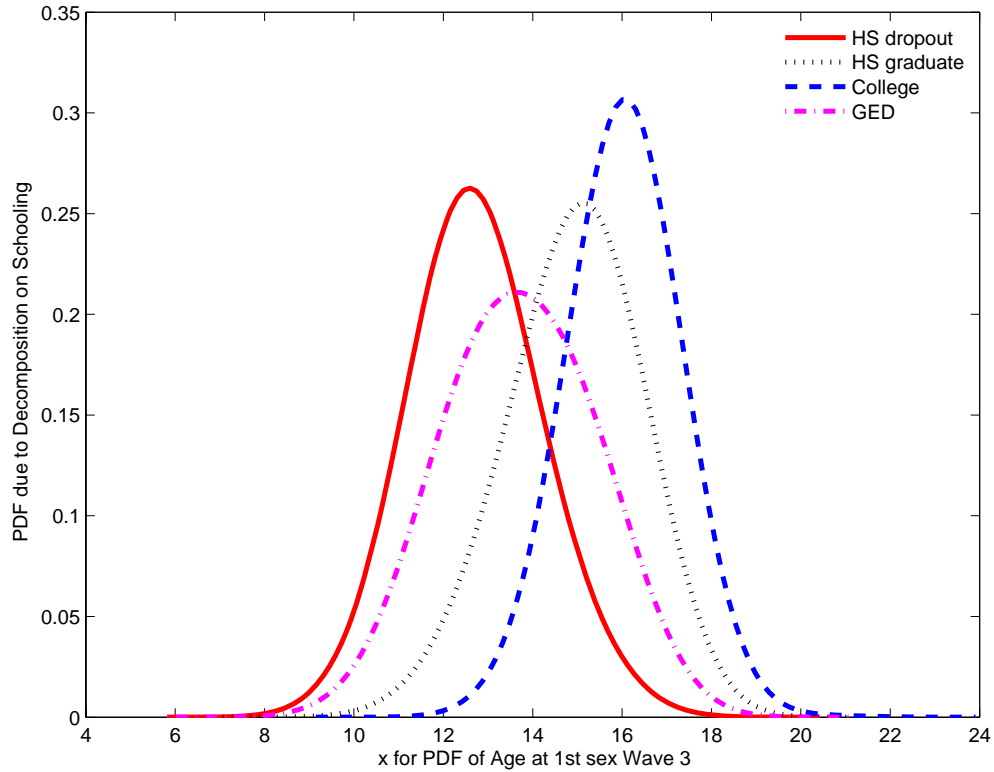**Figure 38:** Delinquency Index , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
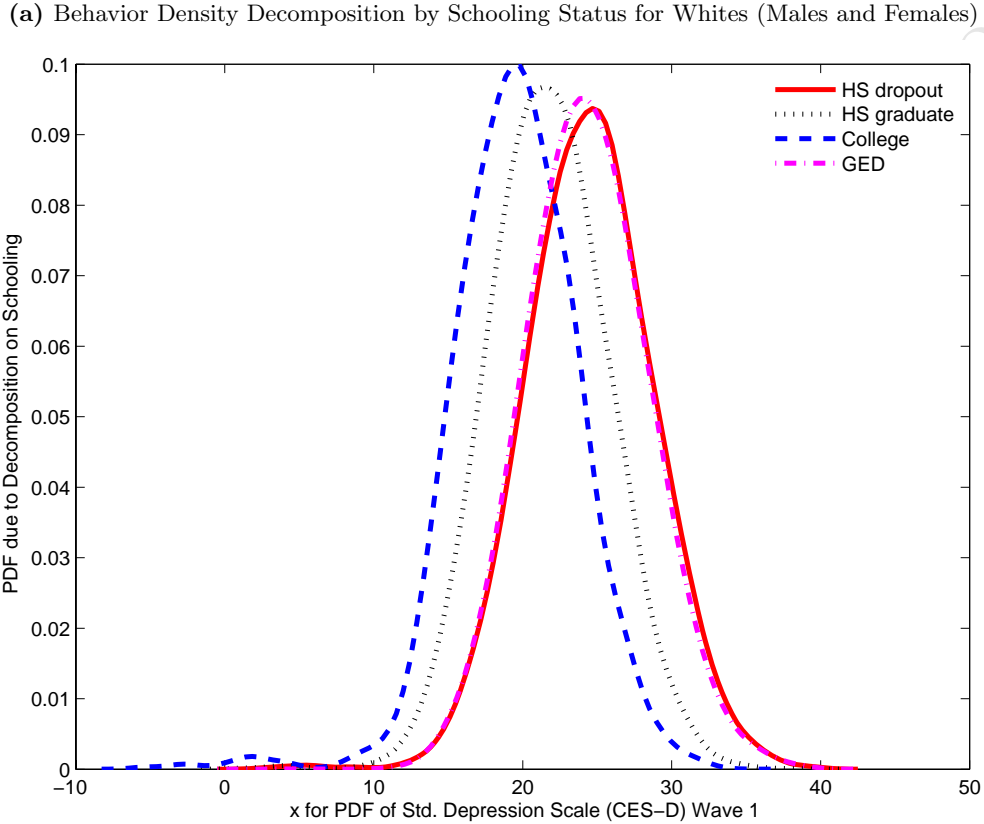
Figure 39: Age at First Sex , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 40: Center for Epidemiological Studies Depression (CES-D) Scale, Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 41: Center for Epidemiological Studies Depression (CES-D) Scale , Wave II
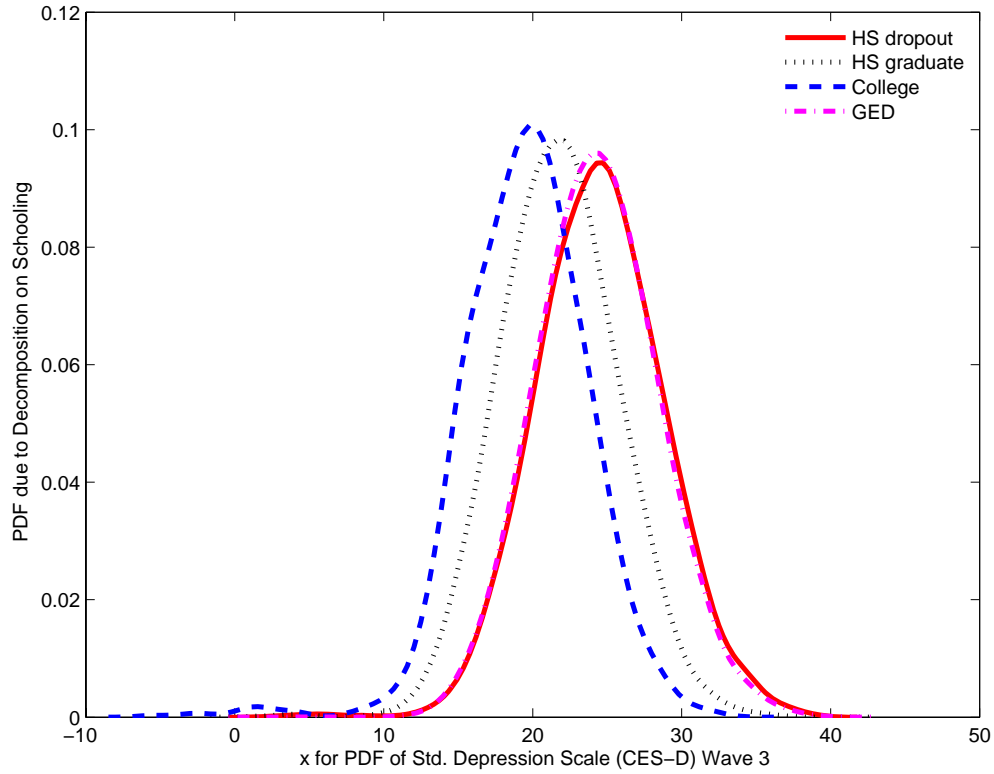
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 42: Center for Epidemiological Studies Depression (CES-D) Scale , Wave III
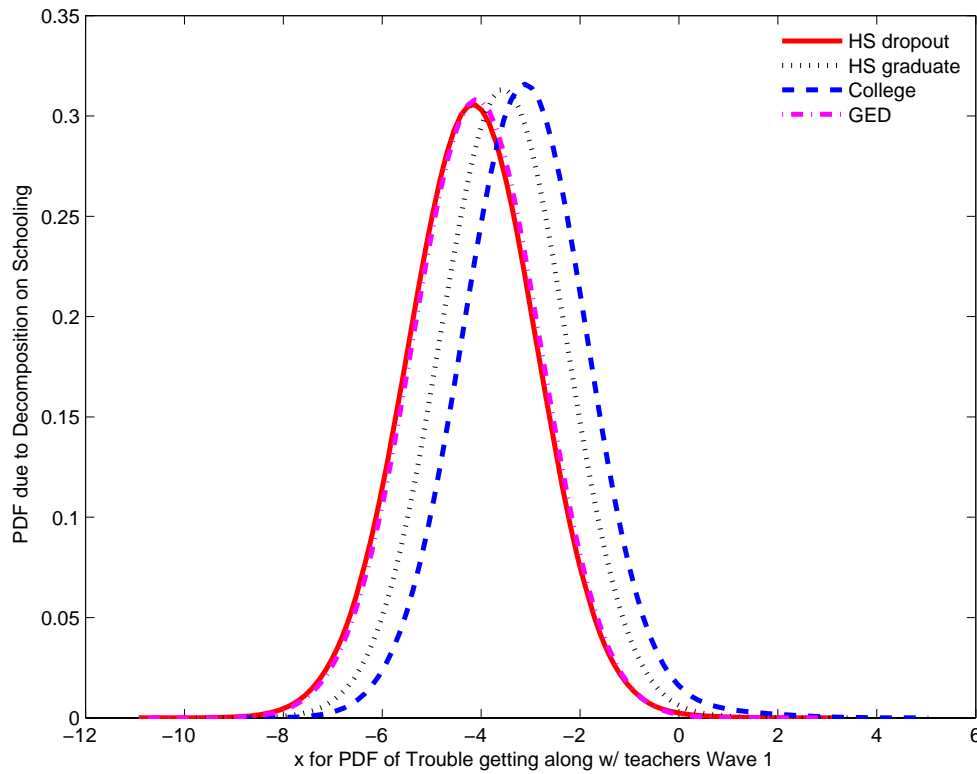
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 43: Have Trouble Getting Along with Teachers in School , Wave I
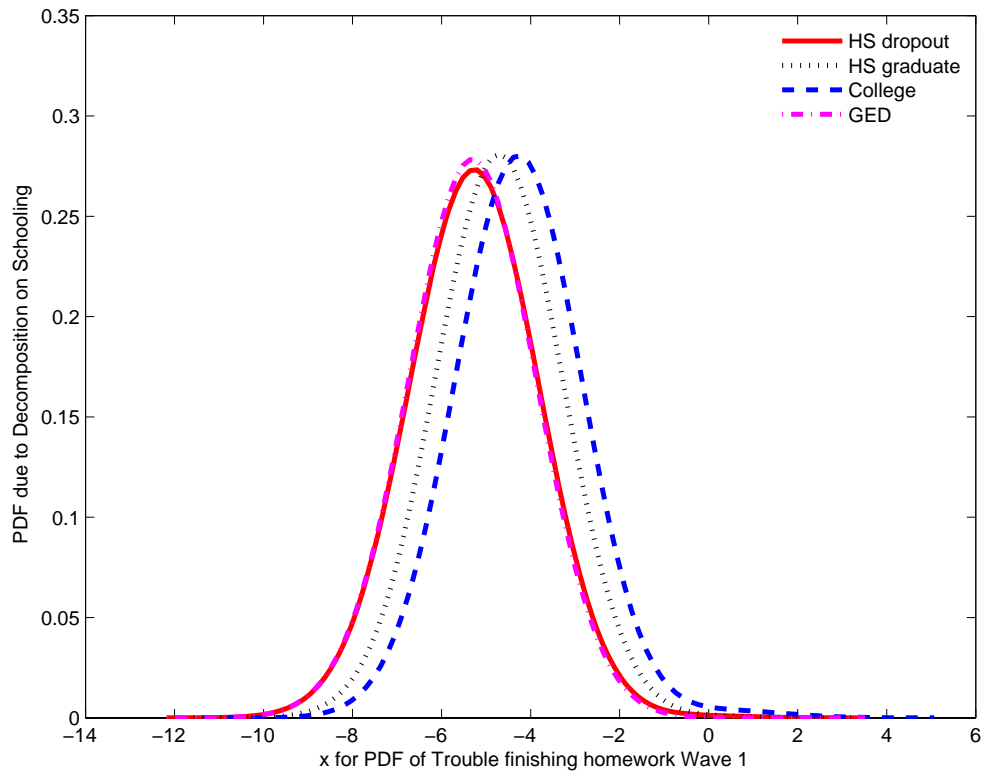
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

49

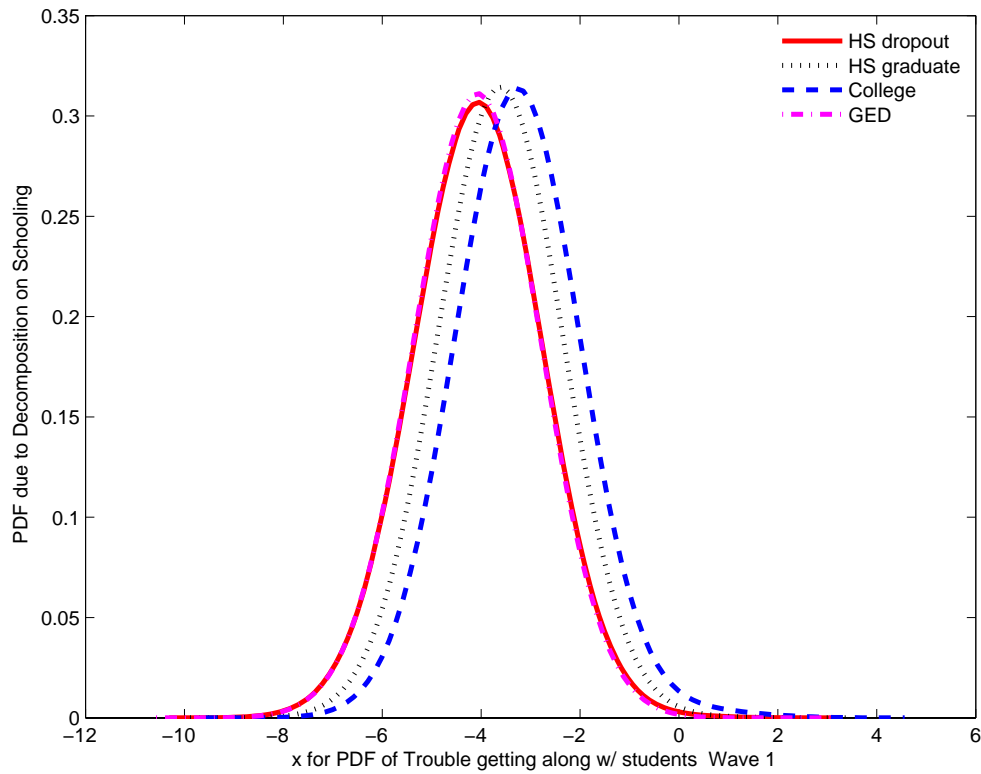**Figure 44:** Have Trouble Finishing Homework , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

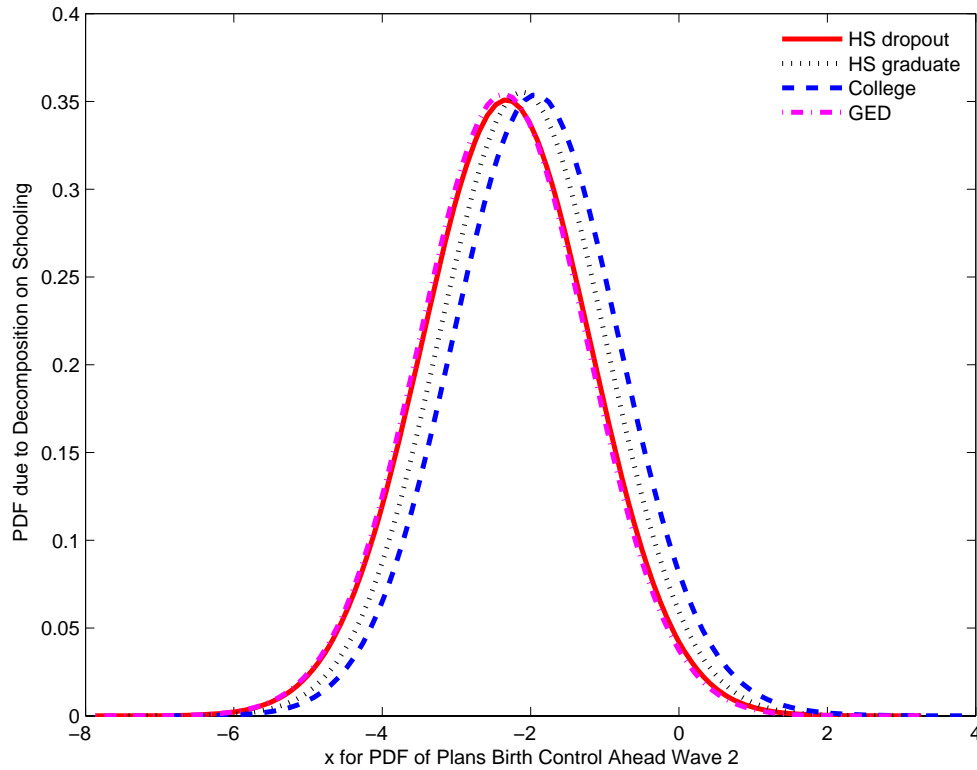**Figure 45:** Have Trouble Getting Along with Students , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

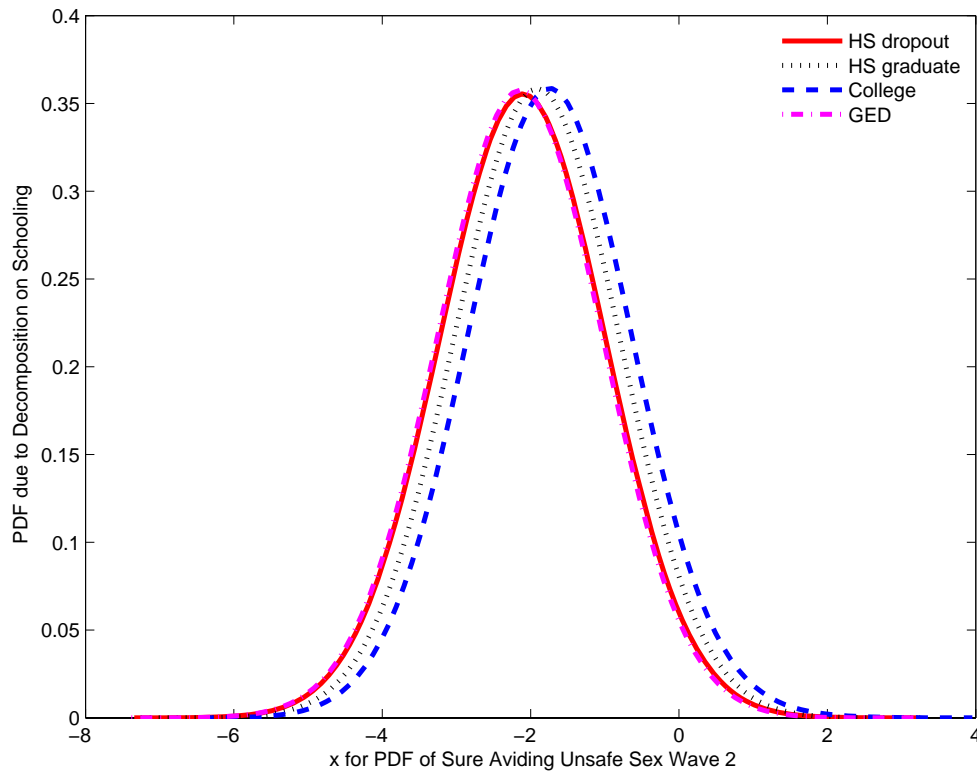**Figure 46:** Plan Birth Control before Sex , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

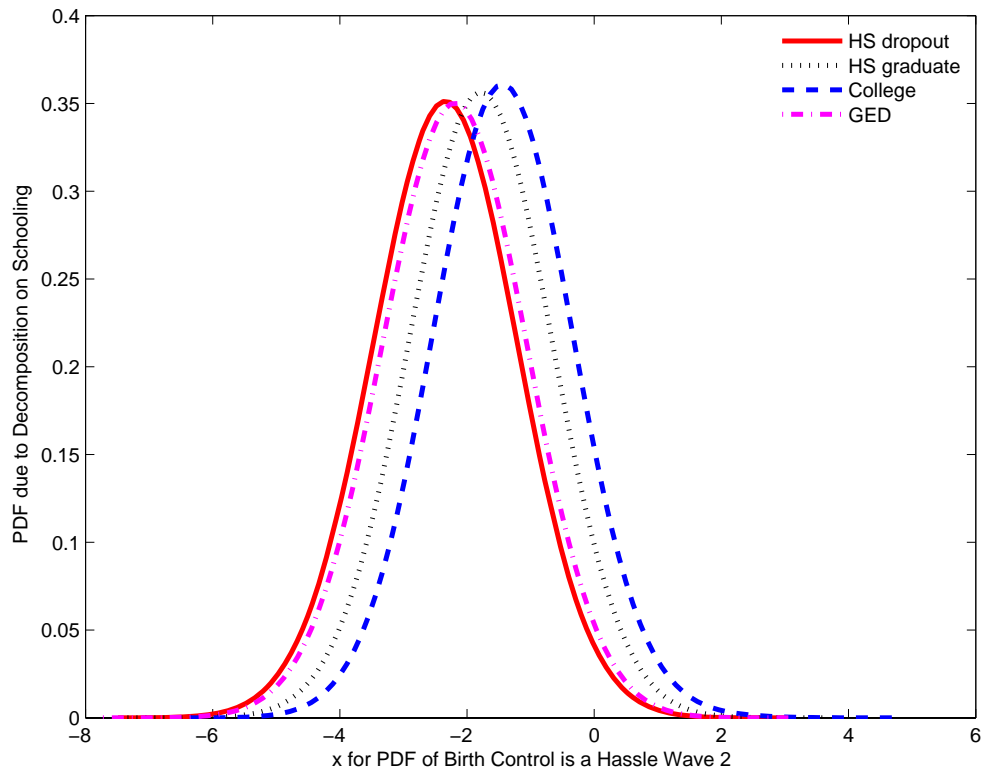**Figure 47:** Sure You Would Avoid Unprotected Sex , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

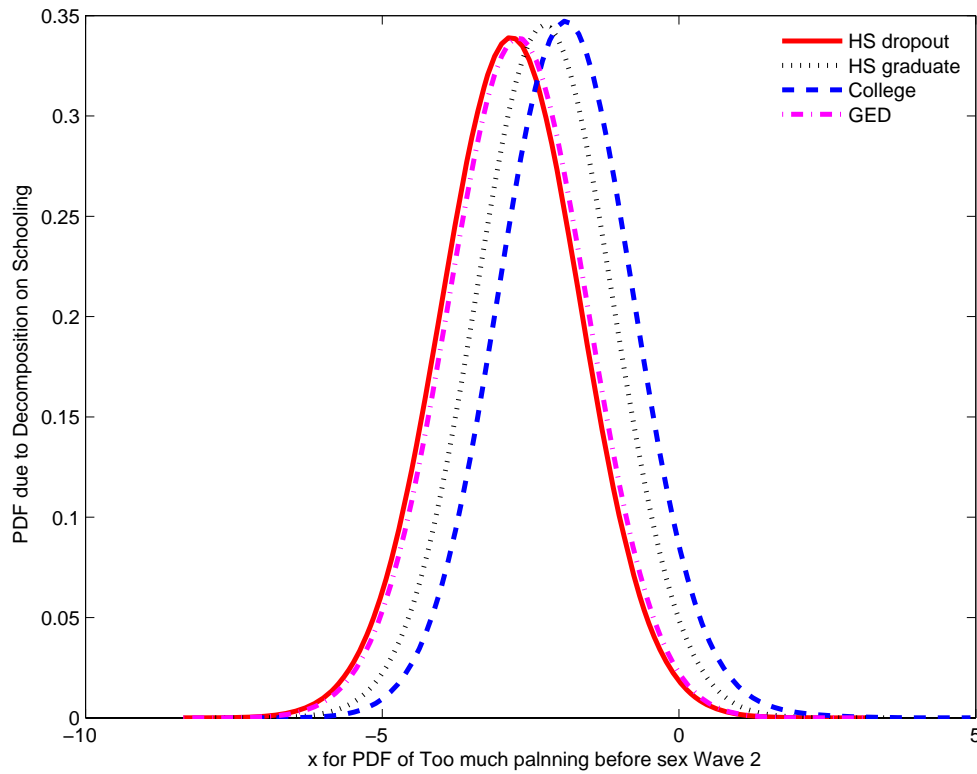**Figure 48:** Birth Control is a Hassle , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 49: It Is Too Much To Plan Birth Control before Sex , Wave II
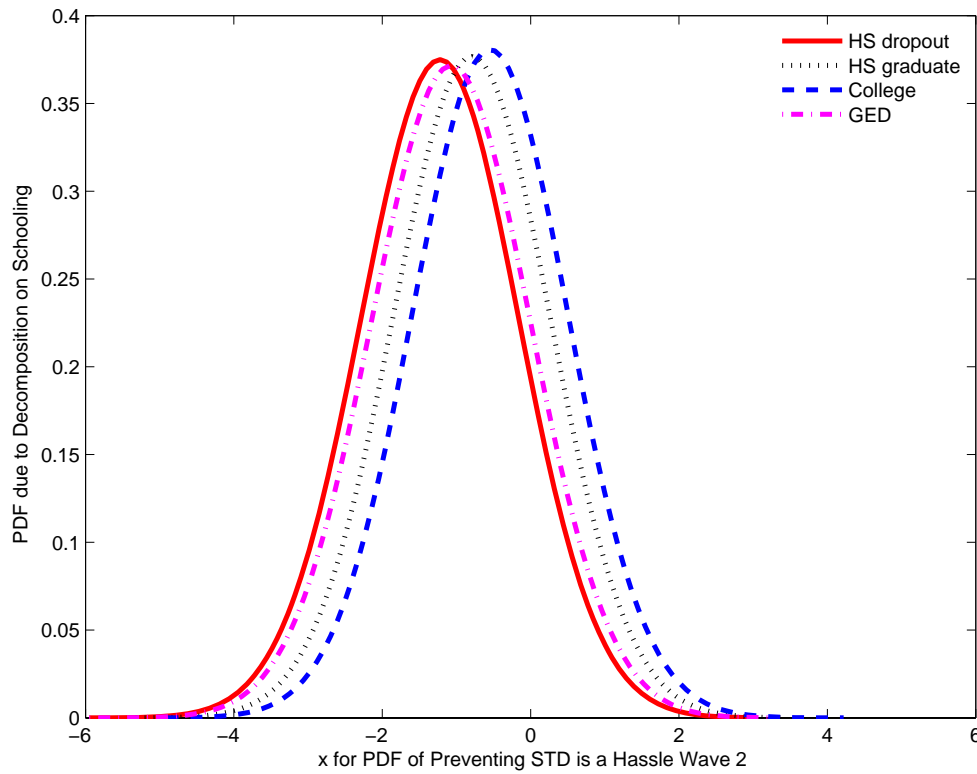
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 50: Preventing STD Is a Hassle , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

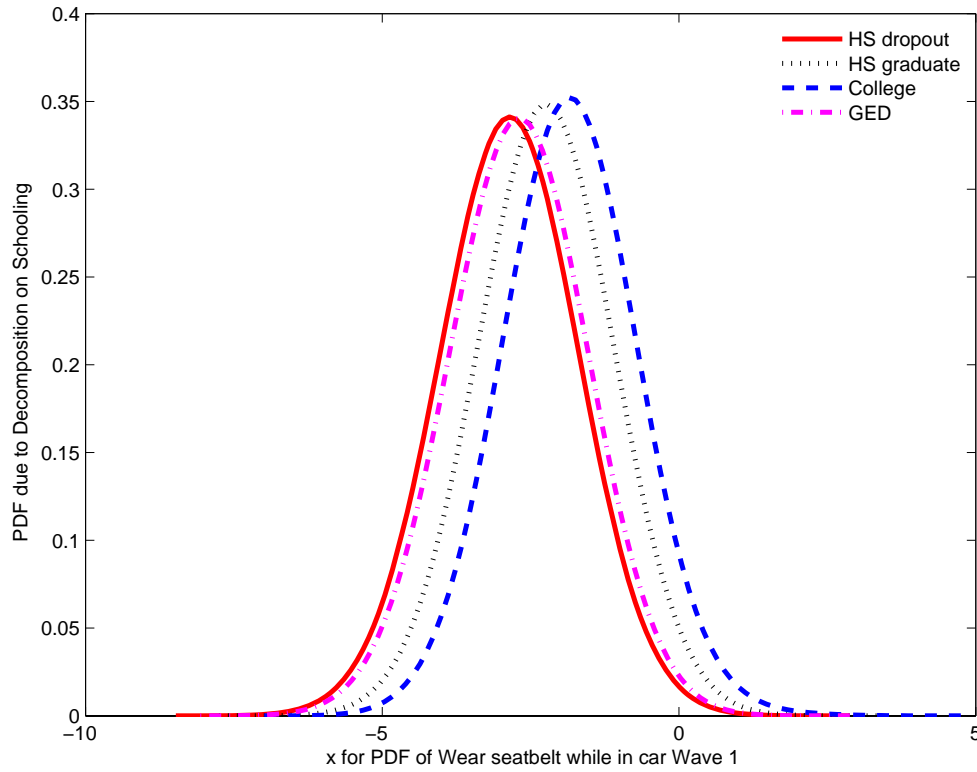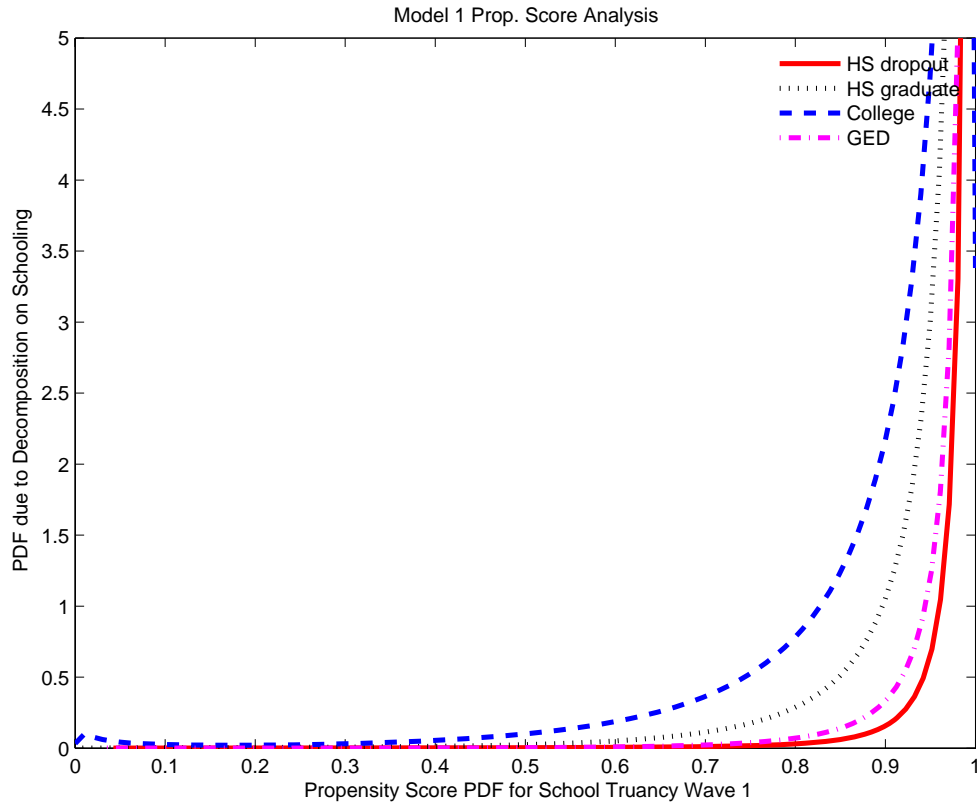**Figure 51:** Wear Seat-belt While Driving or Riding a Car , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 52: School Truancy , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
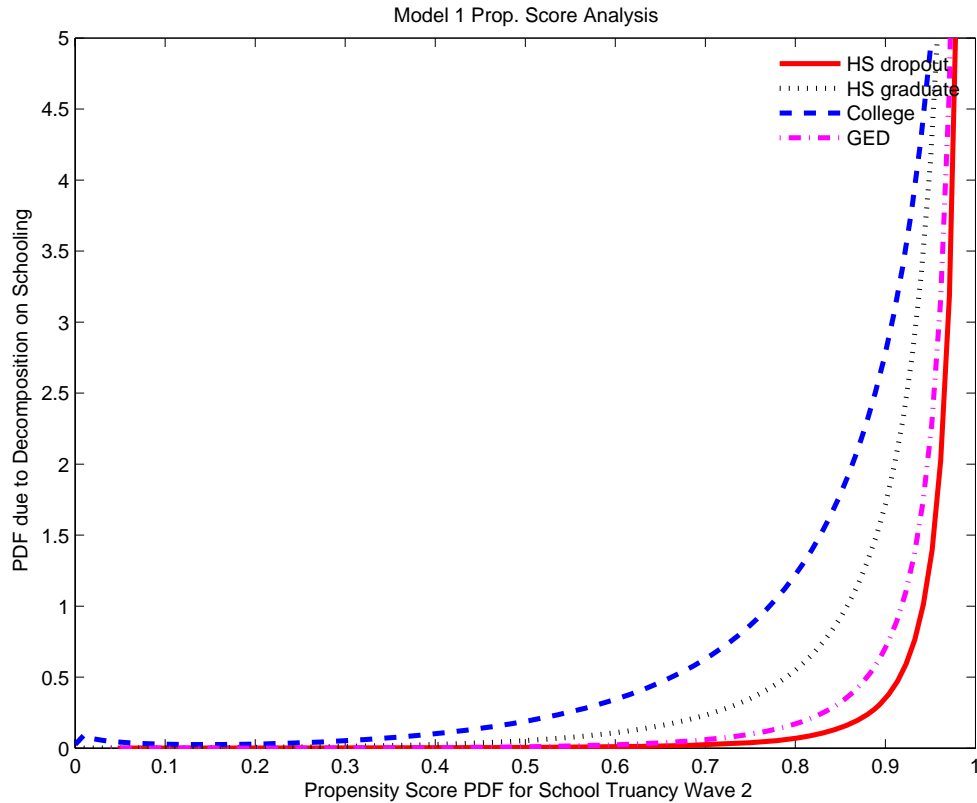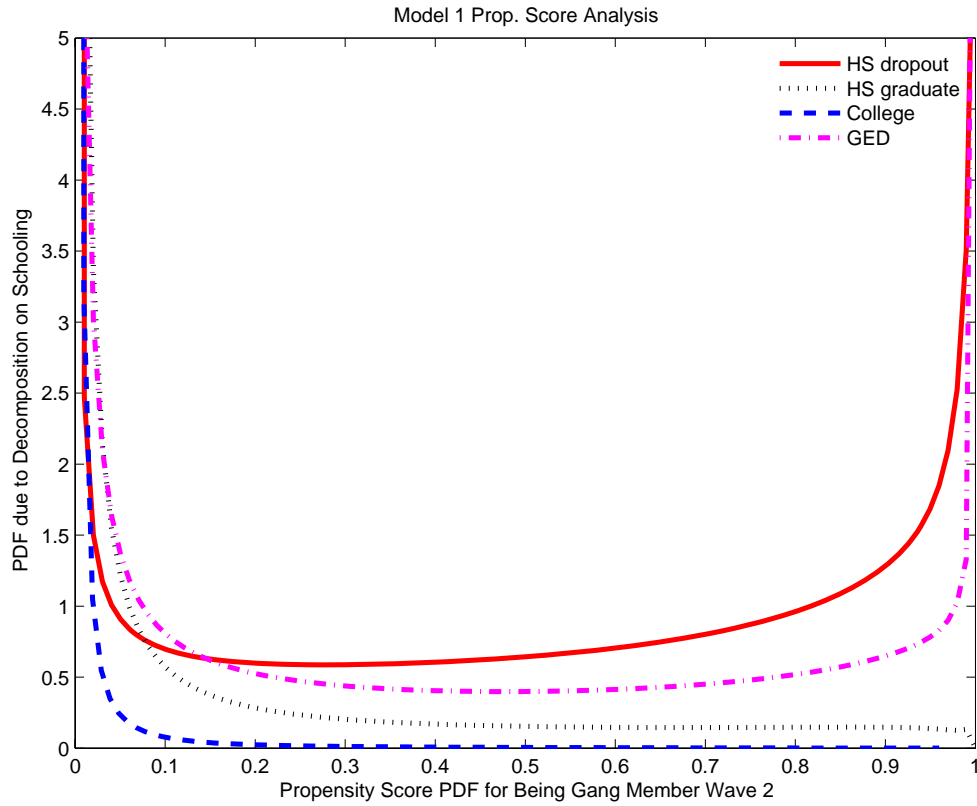
Figure 53: School Truancy , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 54: Even Been a Gang Member , Wave II

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

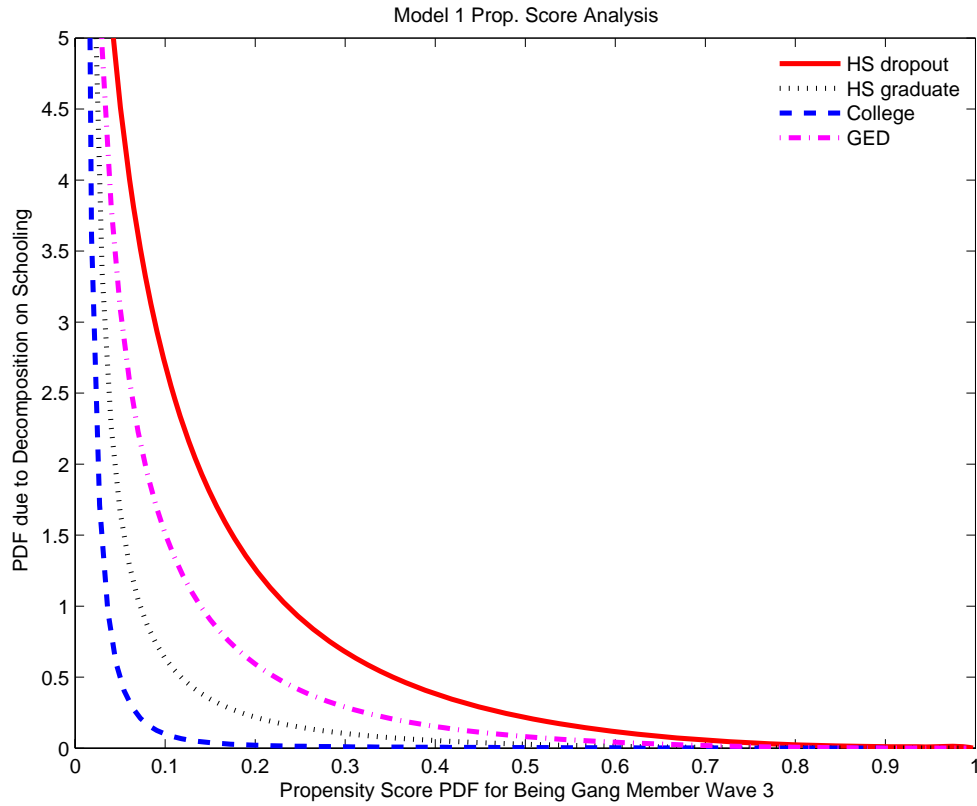**Figure 55:** Even Been a Gang Member , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

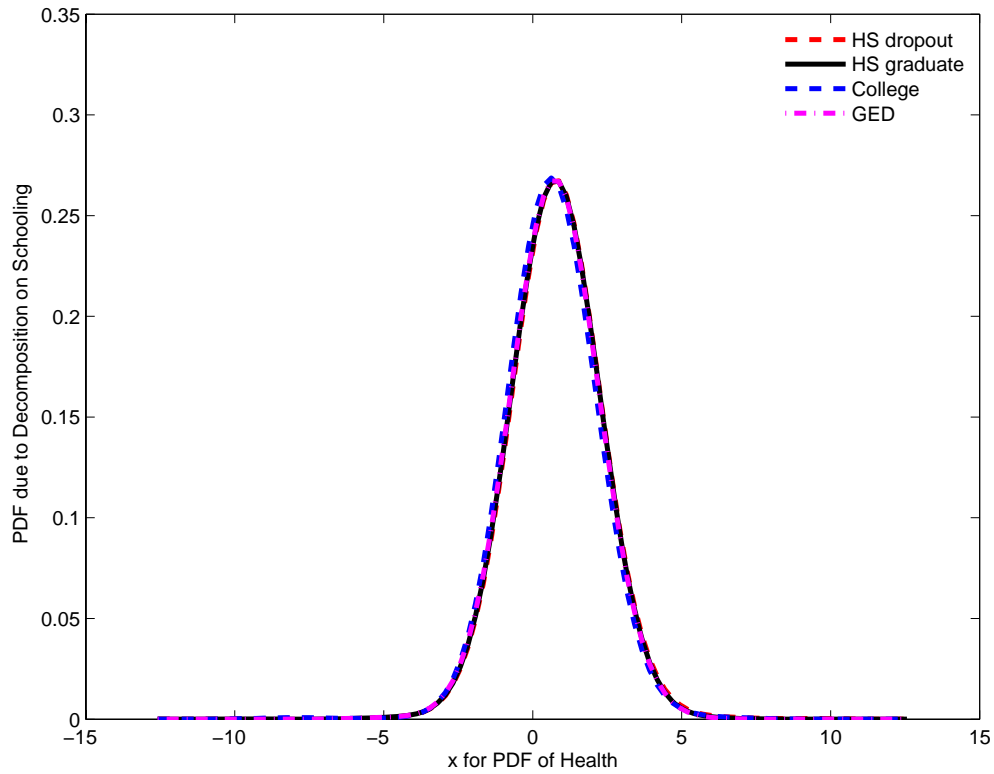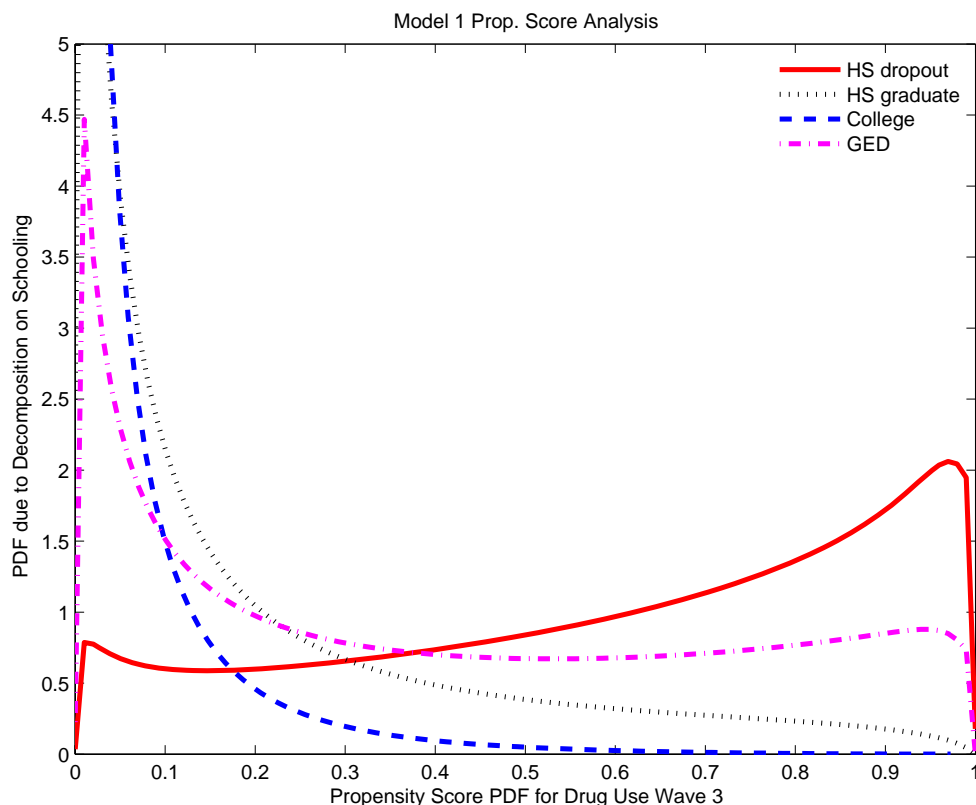**Figure 56:** Health as Latent Factor

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
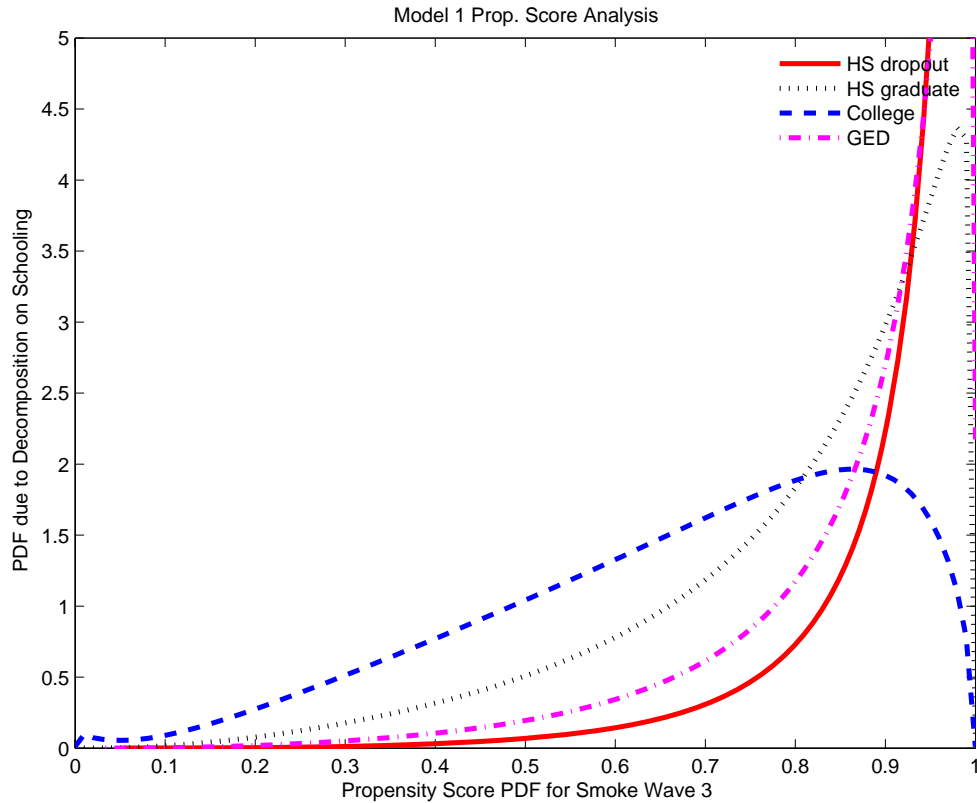
Figure 57: Drug Use , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

63

Figure 58: Smoking , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 59: Ever Arrested , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

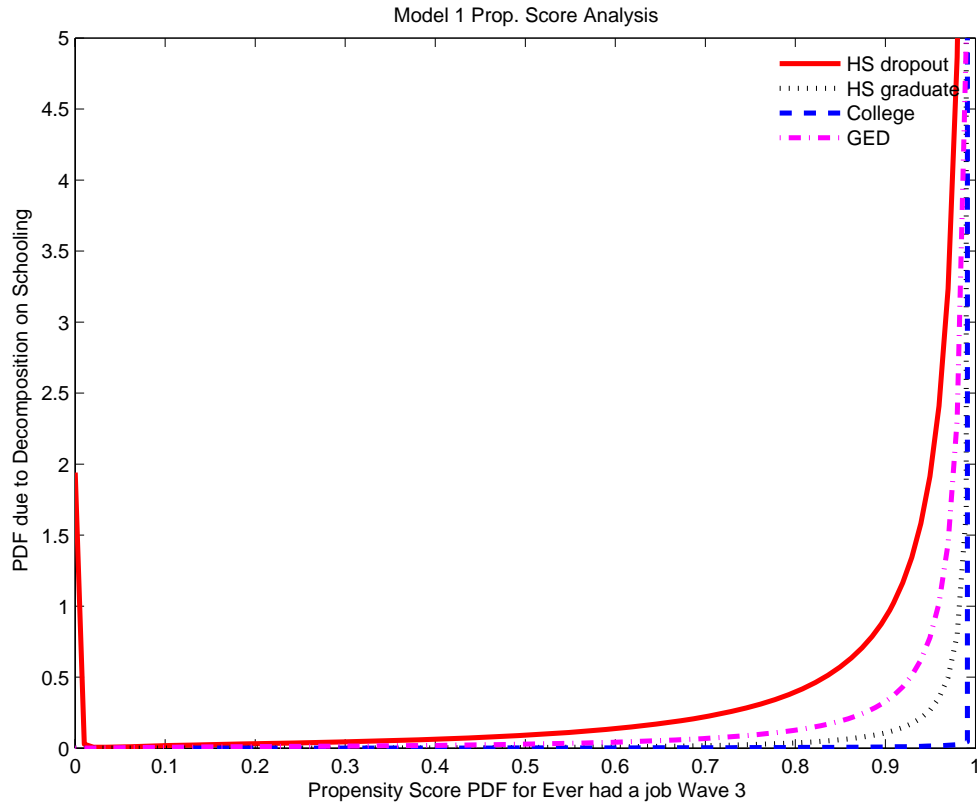**Figure 60:** Ever Had a Job , Wave III
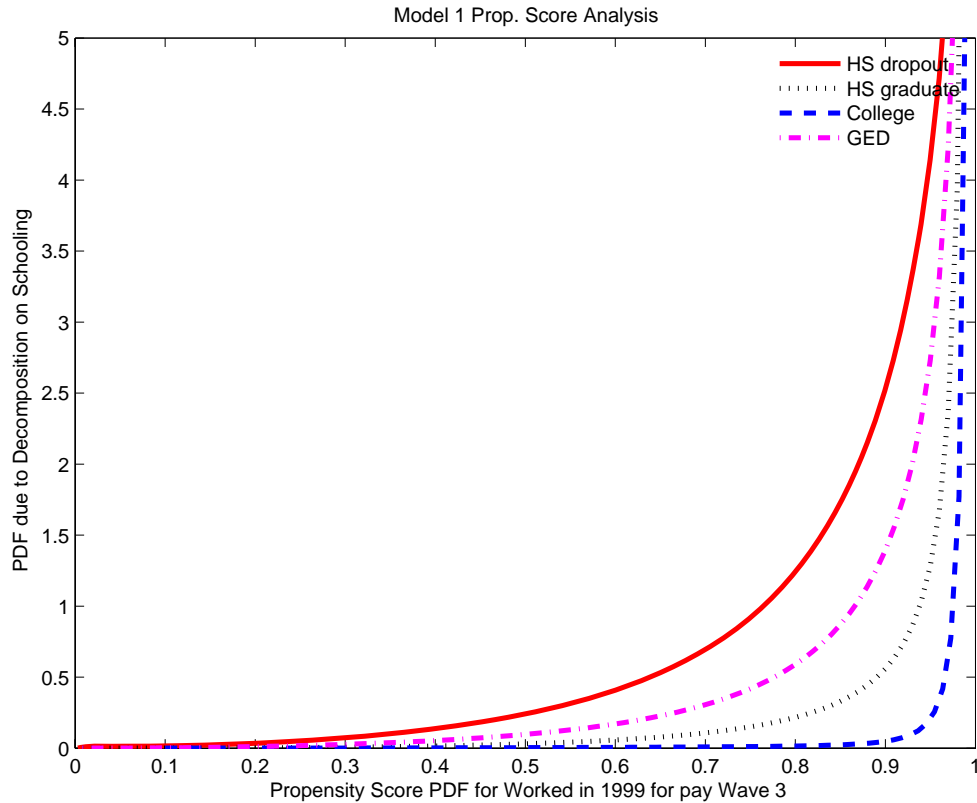
(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 61: Worked for Pay in 1999 , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
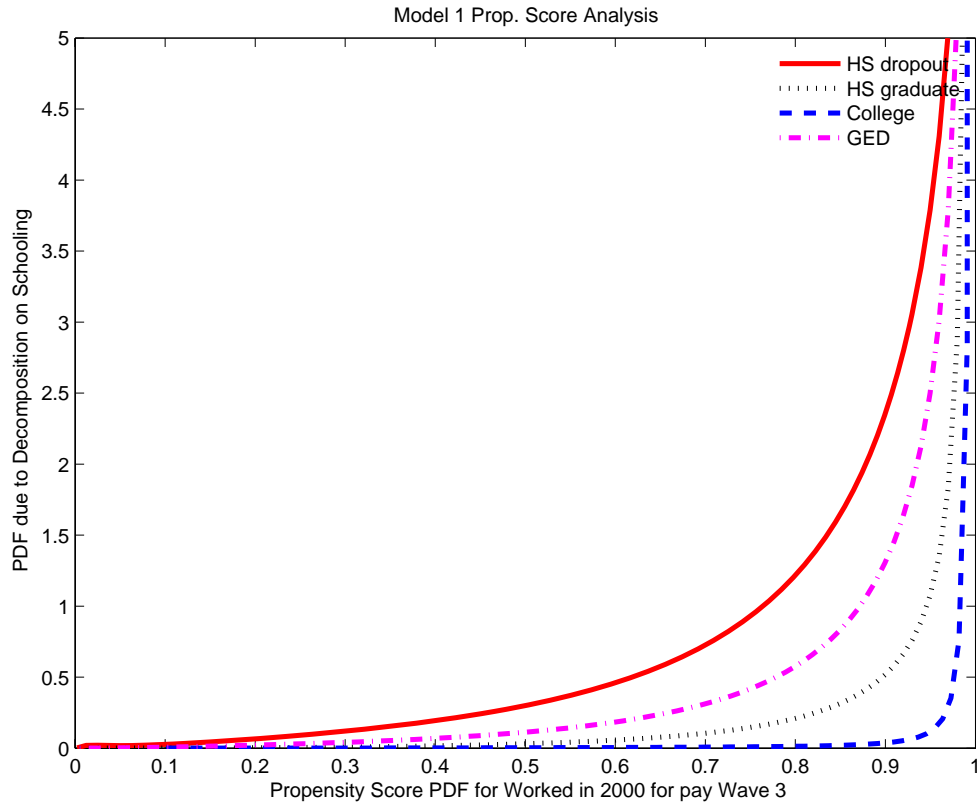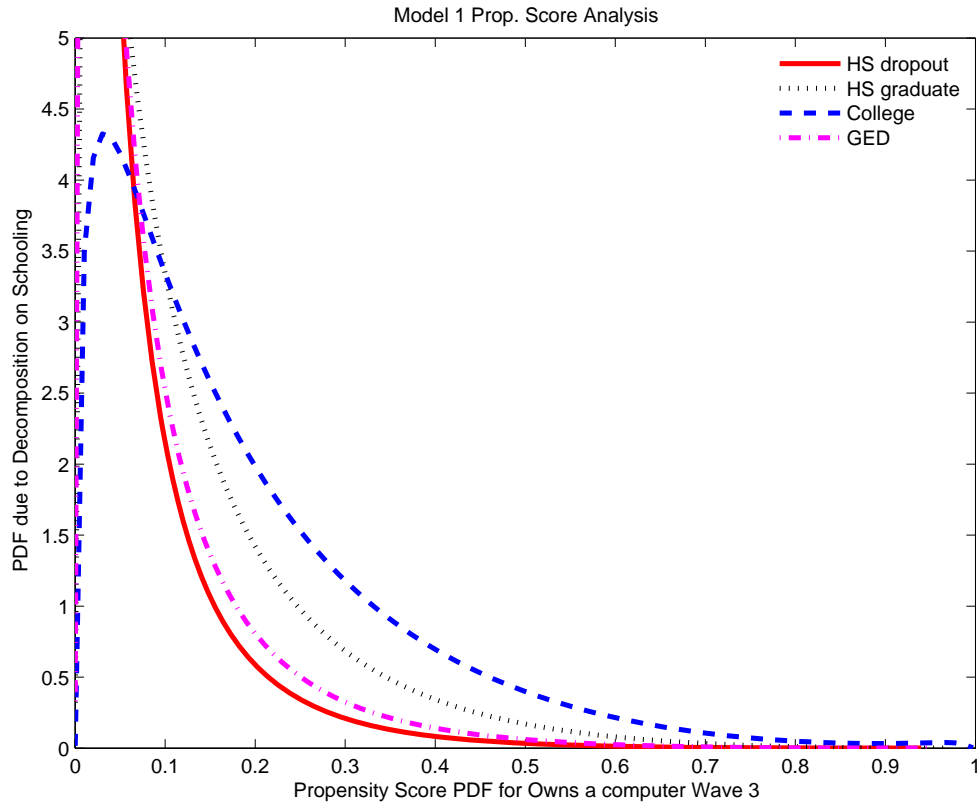
Figure 62: Worked for Pay in 2000 , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
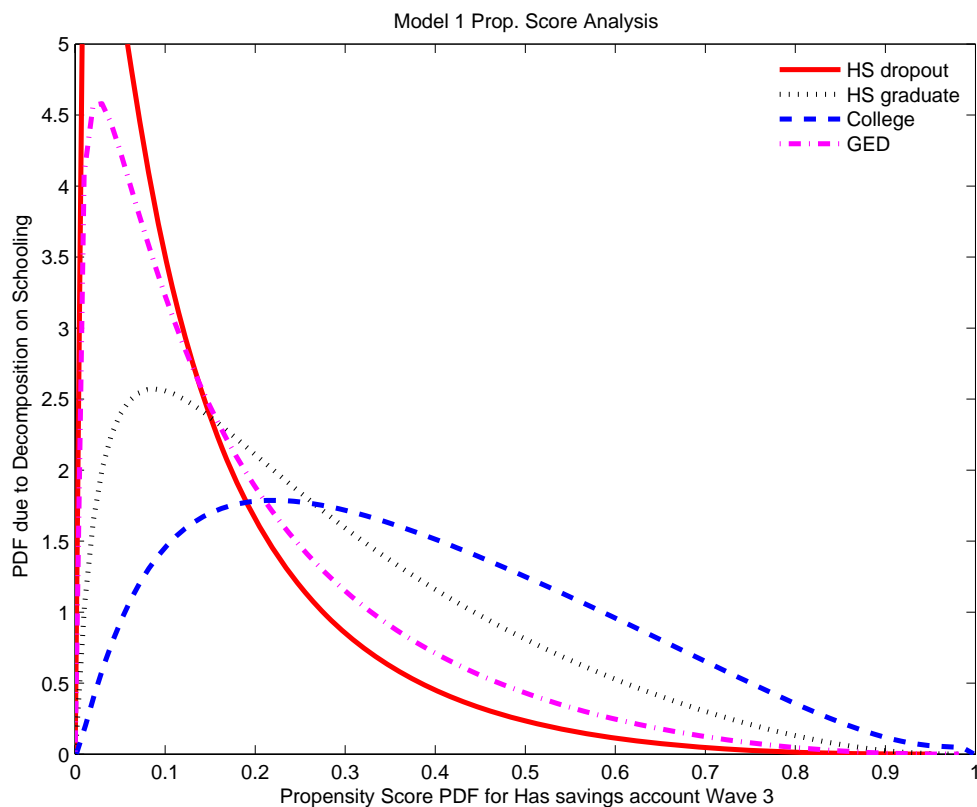
Figure 63: Own a Computer , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
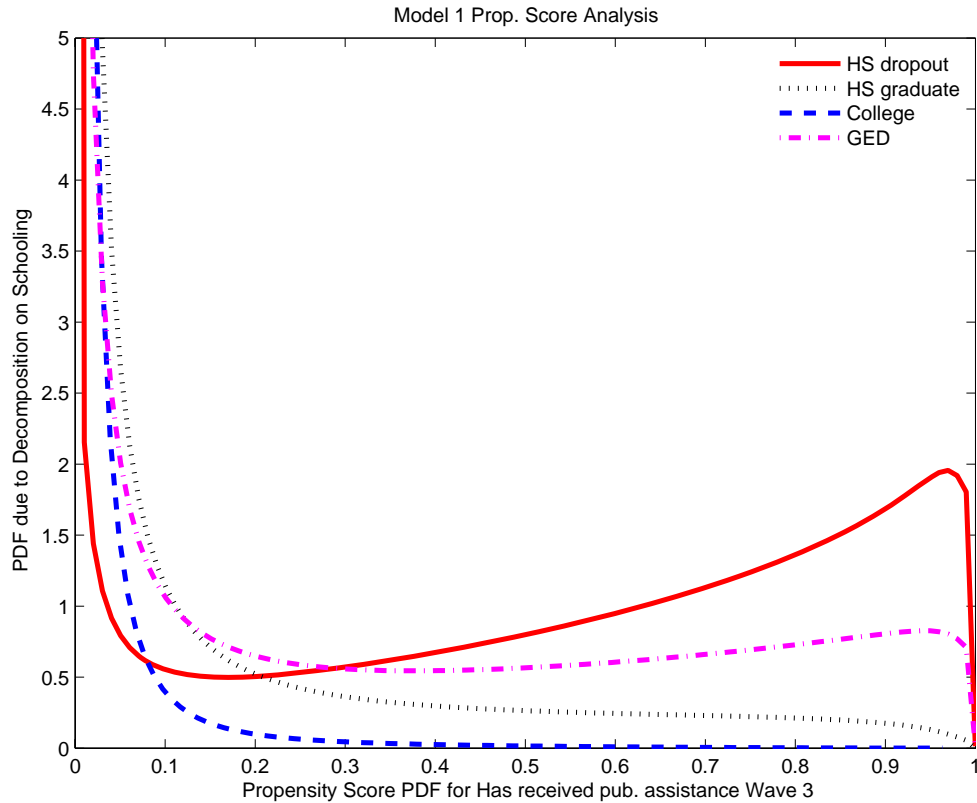
Figure 64: Have a Saving Account , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
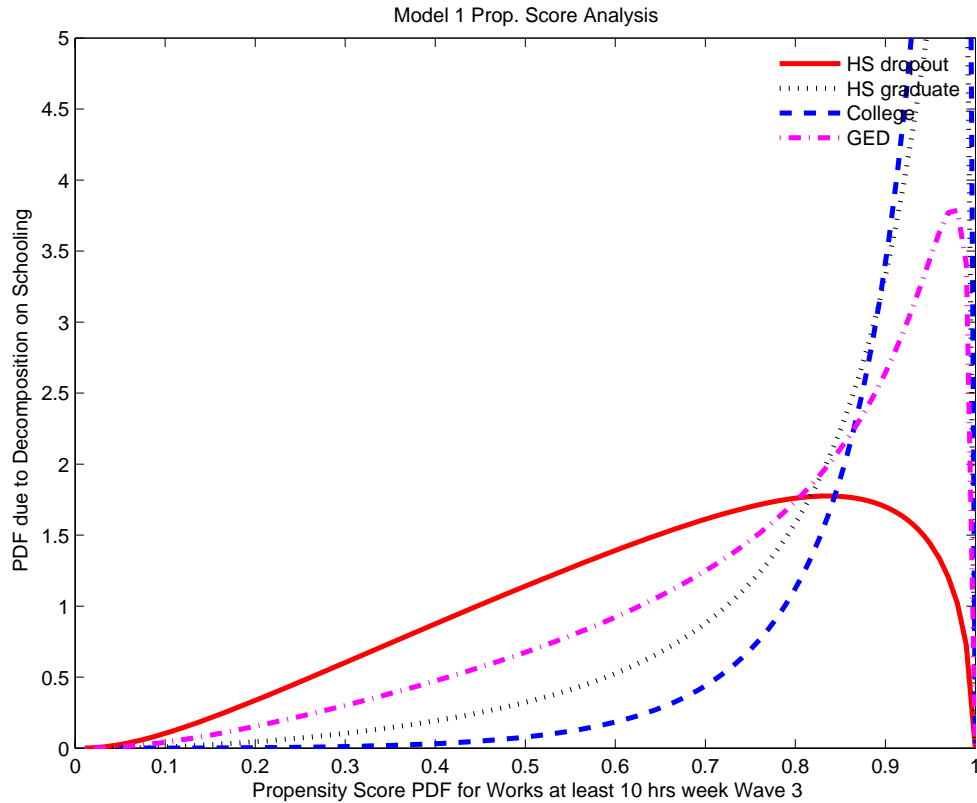
Figure 65: Have Received Public Assistance , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
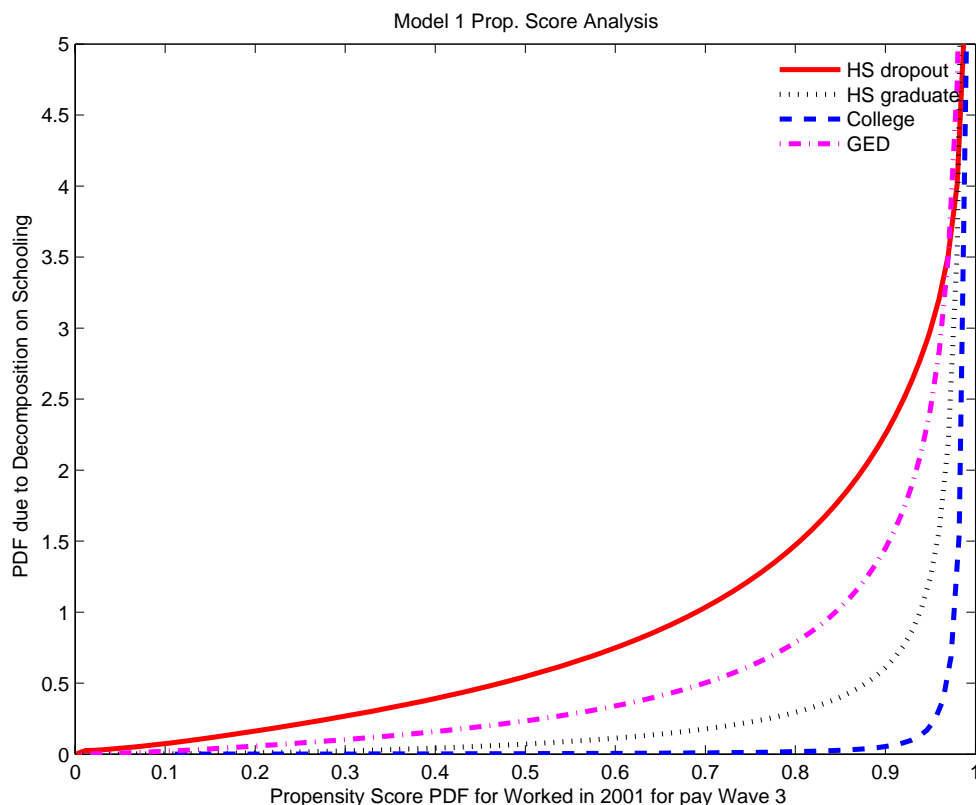
71

Figure 66: Work at Least 10 Hours per Week , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \boldsymbol{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\boldsymbol{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
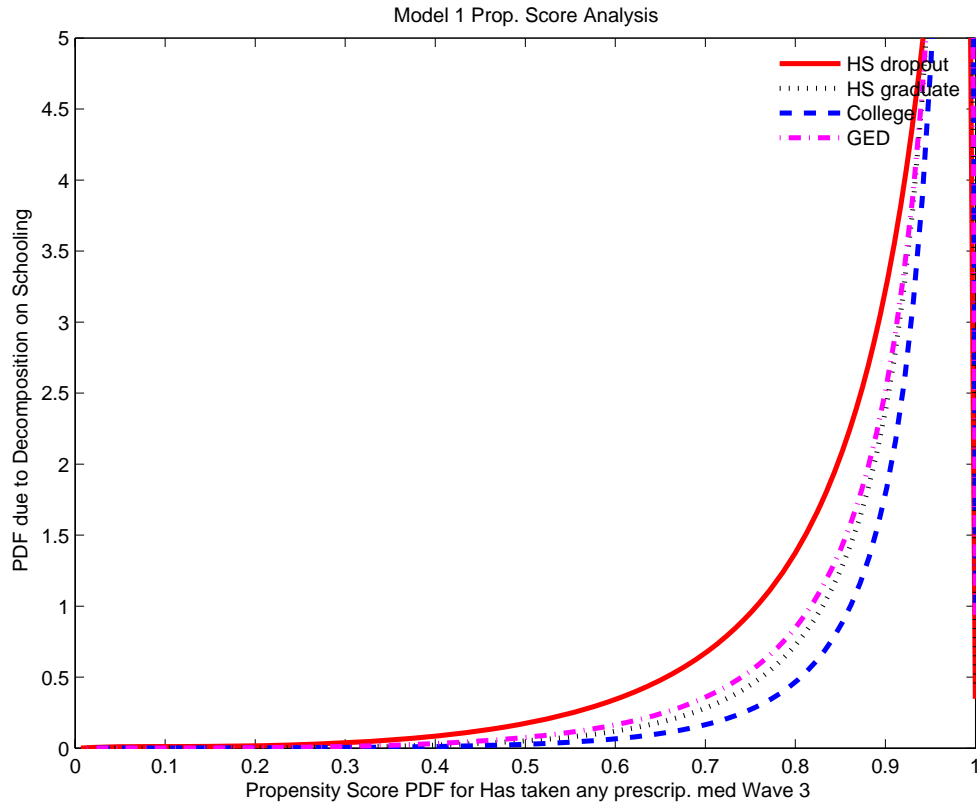
Figure 67: Worked in 2001 for Pay , Wave III

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \mathbf{\Theta}, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, $\mathbf{\Theta}$ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.
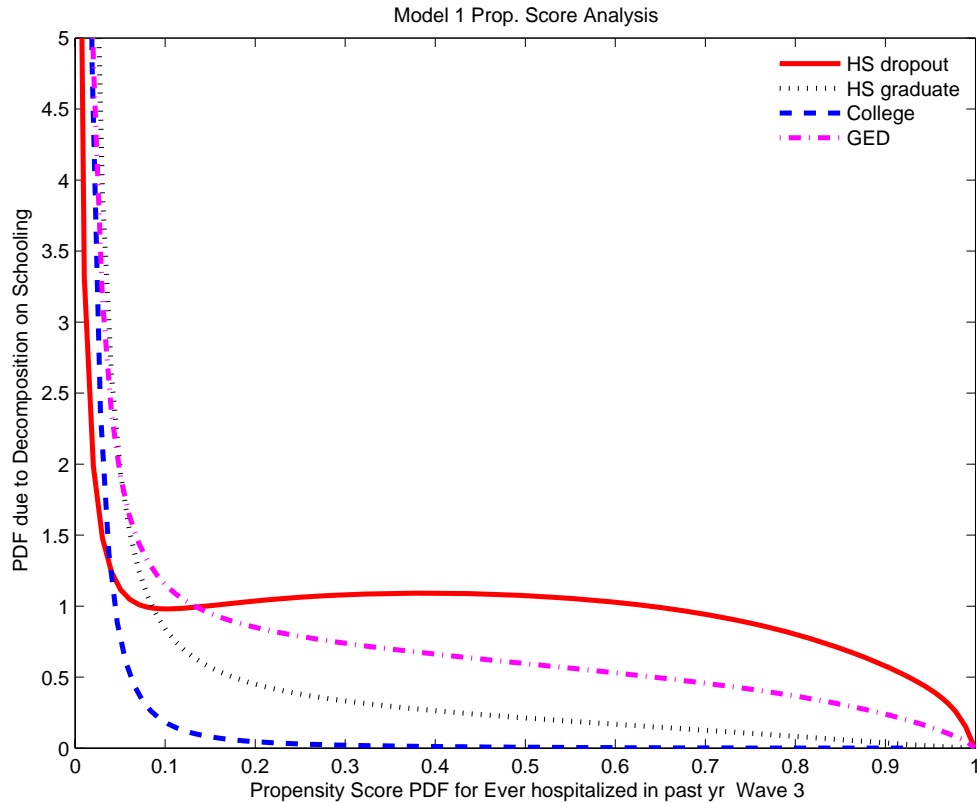
Figure 68: Have Taken Any Prescribed Medicine , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

74

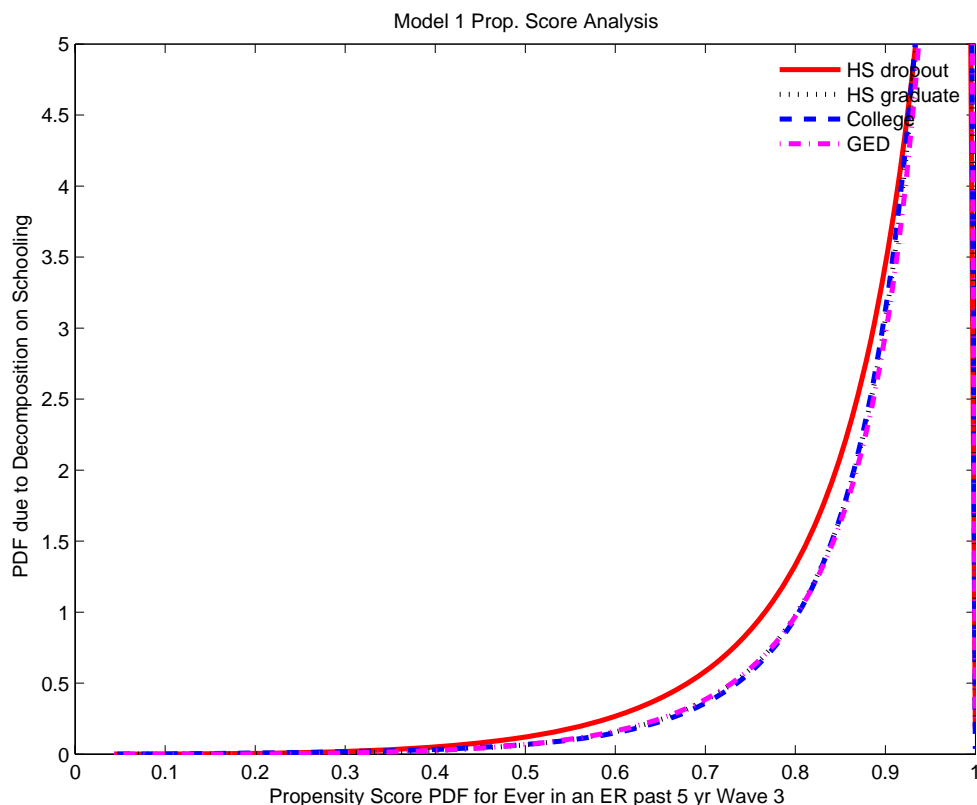**Figure 69:** Ever Been Hospitalized in the Past Year , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

Figure 70: Ever in an ER in the Past 5 Years , Wave I

(a) Behavior Density Decomposition by Schooling Status for Whites (Males and Females)

Model 1 Prop. Score Analysis

Notes: Estimation by Bayesian Latent Factor Model. There are three factors which structural meaning: cognitive skills, noncognitive skills and health condition. Data is classified into measures and outputs. Outputs are loaded into all factors. Measures are loaded into a single output. Outcomes's type are continuous, categorical or binary, thus they are modeled by linear, multinomial and probit models respectively. All measures are continuous. Measures of noncognitive skills are: Peabody Vocabulary Test standardized score, Math GPA and math level at ninth grade. Noncognitive measures are drawn form substantial scales of traits of personality in Add Health Inventory. Health measures are general health status scale and BMI. The estimated equations can be written as $Y_i = f(\beta \cdot X_i + \alpha \cdot \Theta, \epsilon)$, where the function f depends on the outcome type, Y_i is the outcome for participant i, α is the vector of factor loading, Θ is the vector of latent factors. X are pre-program variables such as gender and race, ϵ is an i.i.d. error.

5 Conclusion

Very Preliminary

We presented a model where human skills, such as cognitive, noncognitive and health, are modeled as latent variables. In our model, this skills can affect a large range of lifetime outcomes, such as economic variables, schooling decisions, sexual tendencies and specially risk behavior. We use the Add Health core data set, which is nationally representative, to estimate our model. Finally we, plot the distribution of our results conditional on the schooling choices.

We found that the the cognitive skills increase as we increase the schooling level of the analyzed group. More specifically, the cognitive ability is higher for college people, decreasing for high school graduate, GED graduate and are the high school dropout have the lowest levels of cognitive ability. One interesting result is that the noncognitive ability does not follow the same pattern. In our estimations, the GEDs are the ones with the lowest levels of noncognitive ability. This result corroborates the analysis of Heckman et al. (2006), which state that the GEDs have lower cognitive ability compared to their average cognitive ability.

We also show that schooling, driven by cognitive and noncognitive skills, have an important part in determining the risk behavior. As schooling increases , the probability of becoming a gang member or have an irresponsible sexual behavior decreases. In special, the age of the first intercourse increases with the schooling level. The decomposition of the effect of the skills into each outcome shows often that noncognitive ability plays a major part on determining the levels of many outcomes, in special, the risk behavior outcomes.

References

Borghans, L., B. ter Weel, and B. A. Weinberg (2007, January). Interpersonal styles and labor market outcomes. Working Paper 12846, NBER.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Cunha, F. and J. J. Heckman (2006). Investing in our young people. Unpublished manuscript, University of Chicago, Department of Economics.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*. Forthcoming.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004, July–August). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1-2), 39–98.

Heckman, J. J. (2006, June). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

Heckman, J. J. (2007, August). The economics, technology and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences 104*(3), 13250–13255.

Heckman, J. J. (2008). Schools, skills and synapses. *Economic Inquiry*. Forthcoming.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Ryff, C. D. and B. H. Singer (2005). Social environments and the genetics of aging: Advancing knowledge of protective health mechanisms. *Journals of Gerontology B-Psychological Sciences and Social Sciences 60B*(Special Issue I), 12–23.

W.1 The Role of the Local Economy in Explaining Gender Differences in Treatment Outcomes

W.1.1 Log-likelihood

We assume the ε's are independent normals with mean 0 and standard deviation σ_k^l. Let θ denote the vector of all parameters. This includes the β's, α's and σ's. Since the ε's are independent we can write the likelihood as

$$L(\theta, \boldsymbol{\Theta} | \mathbf{Y}, \mathbf{X}) = f(\mathbf{Y} | \mathbf{X}, \theta, \boldsymbol{\Theta})$$

$$= \prod_{i=1}^{N} \prod_{k=1}^{K_1} \frac{1}{\sigma_k} \phi \left(\frac{Y_{k,i} - \beta_k X_i - \alpha_k \Theta_i}{\sigma_k} \right) \prod_{k=K_1+1}^{K} \sum_{l=1}^{L_k} [Pr(Y_{k,i} = l | X_i, \Theta_i, \theta)]^{\mathbf{1}(Y_{k,i}=l)}$$

where

$$Pr(Y_{k,i} = l | X_i, \Theta_i, \theta) = Pr(Y_{k,i}^l > Y_{k,i}^j, \forall j \neq k | X_i, \Theta_i, \theta)$$

$$= Pr \left(\tilde{\varepsilon}_{\mathbf{k}}^{\mathbf{l}} < \begin{pmatrix} \beta_k^l X_i + \alpha_k^l \Theta_i - \beta_k^1 X_i + \alpha_k^1 \Theta_i \\ \vdots \\ \beta_k^l X_i + \alpha_k^l \Theta_i - \beta_k^{L_k} X_i + \alpha_k^{L_k} \Theta_i \end{pmatrix} \right)$$

$$\tilde{\varepsilon}_{\mathbf{k}}^{\mathbf{l}} = \begin{pmatrix} \varepsilon_k^1 - \varepsilon_k^l \\ \vdots \\ \varepsilon_k^{L_k} - \varepsilon_k^l \end{pmatrix}$$

Notice that this is a vector of length $L_k - 1$. So if we have any multinomial probits we need to know how to compute the multivariate normal cdf numerically. We can write it equivalently as a multidimensional integral

$$Pr(Y_{k,i} = l | X_i, \Theta_i, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{c_1 + \varepsilon_l} \cdots \int_{-\infty}^{c_{L_k} + \varepsilon_l} f(\varepsilon_1, \ldots, \varepsilon_{L_k}) d\varepsilon_l d\varepsilon_1 \ldots d\varepsilon_{L_k}$$

$$c_j = \beta_k^l X_i + \alpha_k^l \Theta_i - \beta_k^j X_i + \alpha_k^j \Theta_i$$

If we no longer treat the factors, $\boldsymbol{\Theta}$ as parameters we can write a different likelihood. Suppose Θ_i is distributed iid across individuals i with pdf $f(\Theta)$. Then the likelihood is

$$L(\theta|\mathbf{Y}, \mathbf{X}) = \int L(\theta, \boldsymbol{\Theta}|\mathbf{Y}, \mathbf{X}) f(\Theta) d\Theta$$