Web Appendix for “Victimization and Skill Accumulation: The Case of School Bullying”

Miguel Sarzosa*
Purdue University

November 3, 2021

*403 W. State Street, West Lafayette, IN, 47907; phone: (765) 494-4343; fax: (765) 494-9658; email, msarzosa@purdue.edu.

Contents

Appendix

Web Appendix

Appendix

A Attrition Analysis

In this Appendix, I present some estimations regarding the observations lost due to attrition. The KYP-JHSP lost 7.5% of the observations to attrition from wave 1 to wave 2. Tables A.1 and A.2 show that there are few differences between those who left the sample and those who stayed. The only observable characteristics in which the attrited and the non-attrited subsamples differ are income, the proportion of fathers with graduate school and two of the cognitive tests. These differences are significant at the 90% confidence level. It is important to note that there are no statistical differences between the subsamples according to bullying perpetration, victimization or non-cognitive skills. Table A.2 analyzes the probability of staying in the sample in terms of observable and unobservable characteristics. It shows that, consistent with the findings in Table A.1, the kids that leave the sample are low cognitive skilled wealthy kids with highly educated parents, all of the characteristics that do not correlate with victimization.

Table A.1: Difference in Observables at $t = 1$ of Attrited and Non-Attrited Observations

Variable	Mean Att	Mean Stay	Diff.	Variable	Mean Stay	Diff.
MOB	8.6346	8.9626	-.328	Biparental	.9294	-.0099
Male	.5019	.5	.0019	Mom Only	.0332	.0051
Older Sib.	.4559	.5452	-.0893*	FatherEd: 2yColl	.0678	.005
Young Sib.	.6398	.6341	.0058	FatherEd: 4yColl	.2974	-.0023
lnInc (pc)	4.5632	4.3275	.2356*	FatherEd: GS	.063	.0711*
Urban	.8659	.8676	-.0017	Locus of Control	-.0052	.0682
Bullied	.2107	.2262	-.0154	Irresponsibility	.0068	-.0895
Bully	.2759	.2437	.0321	Self-Esteem	-.0006	.0074
				Lang & SS	.0074	-.0981
				Math & Sc	.0119	-.1576*
				Yearly Test	.009	-.117*

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. `MOB` stands for Month of birth. `Older Sib.` and `Young Sib.` stand for older and younger siblings. `Lang & SS` stands for Language (Korean) and Social Studies. `Math & Sc` stands for Math and Sciences. `FatherEd` stands for father's education attainment. `FatherEd: GS` takes the value of 1 if the father holds a graduate degree and zero otherwise.

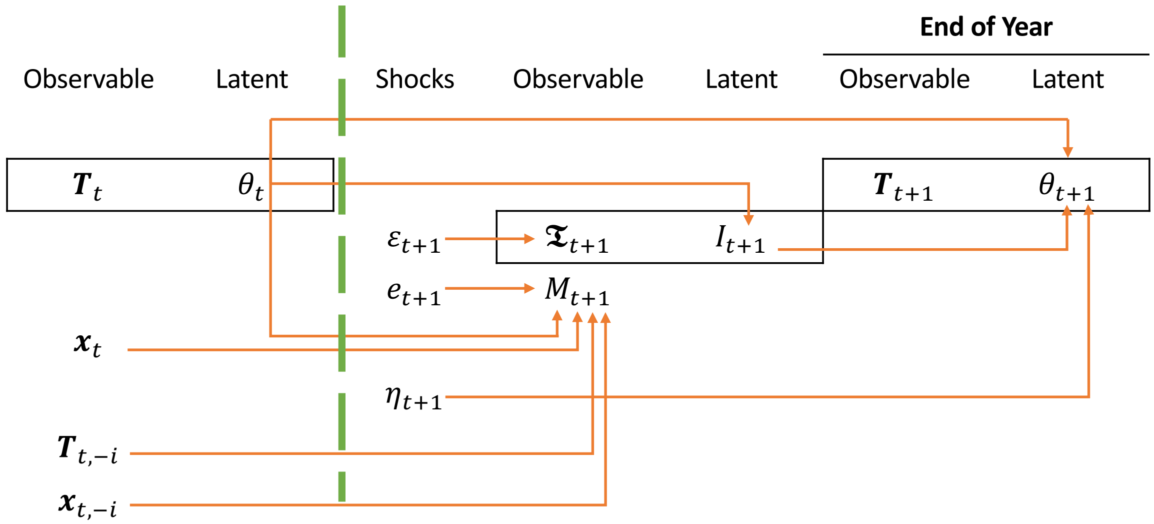
Table A.2: Probability of Staying from $t = 1$ to $t = 2$

Stay in Wave 2	Coeff.	StdErr.
Age (months)	0.0092	(0.010)
Male	-0.0004	(0.072)
Older Siblings	0.0531	(0.070)
Young Siblings	-0.0287	(0.070)
lnInc_pc	-0.3089***	(0.068)
Urban	0.1250	(0.106)
Lives: Both Parents	0.1375	(0.209)
Lives: Only Mother	-0.1876	(0.273)
Father Edu: 2yColl	-0.0036	(0.146)
Father Edu: 4yColl	-0.0727	(0.085)
Father Edu: GS	-0.4410***	(0.126)
Non-Cognitive	-0.2479	(0.321)
Cognitive	0.1497*	(0.078)
Constant	2.5977***	(0.365)
Observations	3,097	

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. `lnInc_pc` stands for log of household income per capita. `Lives: Both Parents` takes the value of 1 if child lives with both parents and zero otherwise. `Lives: Only Mother` takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. `FatherEd` stands for father's education attainment. `FatherEd: GS` takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less.

B Model's Timeline

Figure B.1: Timeline of a Typical Two-period Cycle



Note: Notation is the same used in equations (1)-(6). Each box represents a measurement system that connects a vector of manifest variables with the underlying latent factor (equations (3)-(6)). For the sake of simplicity, in each box, I omit the observable controls that also affect the manifest variables. Also, in the interest of simplicity, I omit the i subindex, but I use the $-i$ subindex to indicate characteristics of peers. $\mathbf{x}_.$ stands for an observable characteristic (e.g., household income per capita). Period t is to the left of the vertical dashed line. Period $t+1$ is to the right of the vertical dashed line. The arrows start from an input and point towards the output. The 'End of Year' columns stress the fact that cognitive scores and non-cognitive measures reflect the end of year endowments.

C Identification of the Model

This appendix presents the identification of the empirical model estimated in this paper. Let me first focus on the identification of $\widehat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$ and $\widehat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$—the estimated latent skills' distributions at t and $t+1$—from (3) and (4). Given the assumptions made, identification of (5) and (6) and, in particular, the latent investment distributions $\widehat{F}_{I_{A,t+1}}(\cdot)$ and $\widehat{F}_{I_{B,t+1}}(\cdot)$ is a special case of the identification problems in (3) and (4). In what follows, I describe identification of $\widehat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$ and the parameters in (3). Identification of $\widehat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ and the parameters in (4) follow the same intuition.

Let $\iota,\iota'=1,\ldots,L$ and $\iota\neq\iota'$ so that $T_{t,\iota}$ represents the ι^{th} manifest measurement at period t. Note that the diagonal elements of the matrix $COV(\mathbf{T}_t\,|\mathbf{X}_{t,T})$ are of the form:

$$COV(T_{t,\iota},T_{t,\iota}\,|\mathbf{X}_{t,T}) = \left(\lambda_t^{T_\iota,A}\right)^2\sigma^2_{\theta_t^A} + \lambda_t^{T_\iota,A}\lambda_t^{T_\iota,B}\sigma_{\theta_t^A\theta_t^B} + \left(\lambda_t^{T_\iota,B}\right)^2\sigma^2_{\theta_t^B} + \sigma^2_{e_t^{T_\iota}} \quad (11)$$

and its off-diagonal elements are of the form:

$$\begin{aligned} &COV(T_{t,\iota},T_{t,\iota'}\,|\mathbf{X}_{t,T}) \\ &\qquad = \lambda_t^{T_\iota,A}\lambda_t^{T_{\iota'},A}\sigma^2_{\theta_t^A} + \left(\lambda_t^{T_\iota,A}\lambda_t^{T_{\iota'},B} + \lambda_t^{T_\iota,B}\lambda_t^{T_{\iota'},A}\right)\sigma_{\theta_t^A\theta_t^B} + \lambda_t^{T_\iota,B}\lambda_t^{T_{\iota'},B}\sigma^2_{\theta_t^B} \end{aligned} \quad (12)$$

where $\lambda_t^{T_{\cdot},\cdot}$ are the elements of Λ_t^T. As it is, the measurement system is underidentified (Carneiro et al., 2003). Assumptions are needed. First, note that latent factors have no metric or scale of their own. This feature poses the need for normalizing to unity one loading per factor. Second, note that loadings, factor variances, and covariances need to be identified from the ${}^{L(L-1)}/_2$ off-diagonal elements of $COV(\mathbf{T}_t\,|\mathbf{X}_{t,T})$ as the diagonal ones will be used to identify $\sigma^2_{e_t^{T_i}}$. Hence, the number of off-diagonal elements needs to be greater or equal to the number of loadings, factor variances, and covariances that will be identified. Given that we are dealing with two factors, this condition implies that I will identify $2L-2$ loadings—due to the normalization of one per factor—two factor variances and one factor covariance. That is, identification requires that the number of manifest measures available is such that the condition $L(L-1)/2 \geq 2L+1$ is fulfilled.[39] Note that this happens if $L \geq 6$.

I follow Carneiro et al. (2003) in assuming that some manifest measures are devoted exclusively to one factor (i.e., assume that $\lambda_t^{T_v,B}=0$ for $v=\{1,2,\ldots,L_A\}$ and $L_A>2$).[40] Therefore, I can organize measurement system (3) such that the subset of measures affected only by θ_A remain on the top L_A rows and the rest of the measures remain in the bottom $L_{A,B}=L-L_A$ rows. That way, I partition the measurement system in two blocks

$$\begin{bmatrix} \mathbf{T}_t^A \\ \mathbf{T}_t^{A,B} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{t,T}\beta^T + \lambda_t^{(A)A}\theta_{A,t} + \mathbf{e}_t^{T^A} \\ \mathbf{X}_{t,T}\beta_t^T + \lambda_t^{(A,B)A}\theta_{A,t} + \lambda_t^{(A,B)B}\theta_{B,t} + \mathbf{e}_t^{T^{A,B}} \end{bmatrix} \tag{13}$$

Then, $COV\left(T_{t,h}^A, T_{t,k}^A \,|\mathbf{X}_T\right) = \lambda_t^{(A_h)A}\lambda_t^{(A_k)A}\sigma^2_{\theta_t^A}$ for $h,k=1,\ldots,L_A$ and $h\neq k$, which yields

$$\frac{COV\left(T_{t,h}^A, T_{t,k}^A \,|\mathbf{X}_T\right)}{COV\left(T_{t,h}^A, T_{t,l}^A \,|\mathbf{X}_T\right)} = \frac{\lambda_t^{(A_k)A}}{\lambda_t^{(A_l)A}}$$

if $h\neq l$ and $l\neq k$. Therefore if, without loss of generality, I normalize the loading of measure l, L_A-1 factor loadings are identified.[41] It is easy to see that once the L_A loadings are identified, $\sigma^2_{\theta_t^A}$ is also identified.

[39] This is slightly different from the assumption required by Carneiro et al. (2003) that specify that $L\geq 5$ for a model with two orthogonal factors. It differs because, as I show below, I depart from the orthogonality assumption between factors and thus, I estimate one additional parameter: the covariance between the factors.

[40] The loading structure of (3) depends entirely on the data available. Ideally, researchers have three measures for each factor, where each measure depends only on one factor. That is, in system (3) we will have the simplest version of Λ_t^T with $\lambda_t^{T_5,A}=0$ and $\lambda_t^{T_6,A}=0$. However, this is not often the case. There are many measures that depend on both latent factors. For instance, grades and education achievement scores may depend not only on a cognitive factor, but also on a non-cognitive one (Heckman et al., 2011).

[41] Given that I use a different set of manifest investment measures for each latent investment factor, identification of (5) and (6) and in particular the latent investment distributions $\widehat{F}_{I_{A,t+1}}(\cdot)$ and $\widehat{F}_{I_{B,t+1}}(\cdot)$ follows the logic used to identify the parameters in the first block of measures in (13).

The second block of the measurement system yields covariance terms of the form:

$$COV\left(T_{t,m}^{A,B}, T_{t,n}^{A,B}\left|\mathbf{X}_{T}\right.\right) = \lambda_t^{(A,B)_m A}\lambda_t^{(A,B)_n A}\sigma^2_{\theta_t^A} + \lambda_t^{(A,B)_m B}\lambda_t^{(A,B)_n B}\sigma^2_{\theta_t^B} + \left(\lambda_t^{(A,B)_m A}\lambda_t^{(A,B)_n B} + \lambda_t^{(A,B)_m B}\lambda_t^{(A,B)_n A}\right)\sigma_{\theta_t^A,\theta_t^B} \quad (14)$$

$$COV\left(T_{t,m}^{A,B}, T_{t,k}^{A}\left|\mathbf{X}_{T}\right.\right) = \lambda_t^{(A,B)_m A}\lambda_t^{(A)_k A}\sigma^2_{\theta_t^A} + \lambda_t^{(A,B)_m B}\lambda_t^{(A)_k A}\sigma_{\theta_t^A,\theta_t^B}$$

for $m, n = L_A+1, \ldots, L_B$, $m \neq n$ and $k = 1, \ldots, L_A$. It is easy to see that the second block of the measurement system is underidentified as it has $2L_B+1$ unknowns, while it has only $L_B(L_B+1)/2$ pieces of relevant information.[42] Therefore if $L_B = 3$, I have seven unknowns and six covariances to use for identification.

This is the reason why one of the main identifying assumptions in Carneiro et al. (2003) is the orthogonality of the factors (i.e., $\theta_A \perp \theta_B$). However, this restriction can only apply to estimations where no factor dynamics are involved. It is easy to see that $\theta_A \perp \theta_B$ cannot be sustained if we believe there are recursive processes governing the production of factor endowments at a given point in time. In particular, in a recursive and intertwined process in which $\theta_{S,t+1} = g_S\left(\theta_{A,t}, \theta_{B,t}\right)$ for $S \in \{A, B\}$, $\theta_{A,t+1} \not\perp \theta_{B,t+1}$ holds because of common past influences. That is, $\theta_{A,t+1}$ and $\theta_{B,t+1}$ are correlated because both share common inputs $\theta_{A,t}$ and $\theta_{B,t}$, even if each latent factor has its own production function $g_A(\cdot,\cdot)$ and $g_B(\cdot,\cdot)$.

In order to allow $\theta_A \not\perp \theta_B$ and still be able to identify the latent factors' distributions and loadings from a measurement system like (3), I propose an additional assumption on the loadings structure Λ_t^T: assume that there is one measure among $\mathbf{T}_t^{A,B}$ that is exclusively affected by the second factor (i.e., $\lambda_t^{(A,B)_o A} = 0$ for $o \in [L_A+1, L_B]$). For presentation simplicity let $T_{t,o}^{A,B}$ also contain the normalized

[42] Unknowns: two loadings per measure minus one that is normalized, $\sigma^2_{\theta^B}$ and $\sigma_{\theta^A,\theta^B}$. Measurement system covariances: $L_B(L_B-1)/2$ covariances within the second block measures and L_B covariances resulting from one covariance between each second block measure and one measure in the first block—preferably, the one that has the normalized loading.

loading for the second factor (i.e., $\lambda_t^{(A,B)_o B} = 1$). Then,

$$
\begin{aligned}
COV\left(T_{t,l}^{A}, T_{t,o}^{A,B} \,|\mathbf{X}_T\right) &= \sigma_{\theta_t^A,\theta_t^B} \\
COV\left(T_{t,m}^{A,B}, T_{t,o}^{A,B} \,|\mathbf{X}_T\right) &= \lambda_t^{(A,B)_m B}\sigma_{\theta_t^B}^2 + \lambda_t^{(A,B)_m A}\sigma_{\theta_t^A,\theta_t^B} \qquad (15) \\
COV\left(T_{t,m}^{A,B}, T_{t,l}^{A} \,|\mathbf{X}_T\right) &= \lambda_t^{(A,B)_m A}\sigma_{\theta_t^A}^2 + \lambda_t^{(A,B)_m B}\sigma_{\theta_t^A,\theta_t^B} \qquad (16)
\end{aligned}
$$

for $m = L_A + 1, \ldots, L_B - 1$ and $m \neq o$. Using (16), I can write $\lambda_t^{(A,B)_m A}$ as a function of $\lambda_t^{(A,B)_m B}$ and together with (15), I can write $\sigma_{\theta_t^B}^2$ as a function of $\lambda_t^{(A,B)_m B}$, which can be replaced in the expression for $COV\left(T_{t,n}^{A,B}, T_{t,o}^{A,B} \,|\mathbf{X}_T\right)$, for $n = L_A + 1, \ldots, L_B - 1$ and $n \neq m, n \neq o$, leaving $\lambda_t^{(A,B)_n B}$ as a function of $\lambda_t^{(A,B)_m B}$ that can be then replaced in (14) to solve the entire system. Having identified all the loadings, $\sigma_{\theta_t^A}^2$, $\sigma_{\theta_t^B}^2$, $\sigma_{\theta_t^A,\theta_t^B}$ and measurement residual variances, together with the fact that the means of θ_A, θ_B and $\mathbf{e^T}$ are finite—something I will return to in Subsection 4.2.2—I use the Kotlarski Theorem to non-parametrically identify the distribution of $f_{\theta_A,\theta_B}(\cdot)$ from the manifest variables $\mathbf{T}_t$ (Kotlarski, 1967).[43]

[43]The Kotlarski Theorem states that if there are three independent random variables e_{T_1}, e_{T_2} and θ and define $T_1 = \theta + e_{T_1}$ and $T_2 = \theta + e_{T_2}$, the joint distribution of (T_1,T_2) determines the distributions of e_{T_1}, e_{T_2} and θ, up to one normalization. Note that, given that we have already identified all the loadings, we can write (3) in terms of $\ddot{T}_\tau = \theta + e_{T_\tau}$—where $\ddot{T}_\tau = T_\tau - \mathbf{X}_{\tau,T}\beta_\tau^T$—by dividing both sides by the loadings. See more details in Carneiro et al. (2003).

D Identification of the CES Function

Table D.1: Relation Between the Mean of $\widehat{\theta}_{t+1}$ and the CES Parameters

	Coef	StdErr		Coef	StdErr		Coef	StdErr
Cons	-0.003***	(0.000)	ρ^3	0.017***	(0.001)	$\gamma_1^2\rho^3$	0.116***	(0.003)
ρ	0.058***	(0.002)	$\gamma_1\rho$	0.722***	(0.008)	$\gamma_2\rho$	0.732***	(0.008)
γ_1	0.006***	(0.001)	$\gamma_1^2\rho$	-0.744***	(0.007)	$\gamma_2^2\rho$	-0.754***	(0.007)
γ_2	0.025***	(0.001)	$\gamma_1\rho^3$	-0.116***	(0.003)	$\gamma_2\rho^3$	-0.117***	(0.003)
$\gamma_2^2\rho^3$	0.117***	(0.003)	$\gamma_1\gamma_2\rho$	-0.476***	(0.019)	$\gamma_1\gamma_2\rho^3$	0.118***	(0.004)
$\gamma_1^2\gamma_2^2\rho$	-1.028***	(0.048)						
Observations		1,440		R^2	0.9998			

*** p<0.01, ** p<0.05, * p<0.1. The $\widehat{\theta}_{t+1}$ plotted are the results of 1,440 different combinations of γ_1, γ_2 and ρ parameters in the CES production function $\widehat{\theta}_{t+1} = [\gamma_1 x^\rho + \gamma_2 y^\rho + (1 - \gamma_1 - \gamma_2) z^\rho]^{1/\rho}$, where x, y and z come from 5,000 random draws from independent normal distributions.

Table D.1 presents the estimates of the regression of the mean of $\widehat{\theta}_{t+1}$ on a cubic polynomial of the parameters of the CES function. These estimates show that the relation between the mean of $\widehat{\theta}_{t+1}$ and the CES parameters presented in Figure 1 is very predictable as the cubic polynomial accounts for 99.98% of the variation of $\widehat{\theta}_{t+1}$.

E Estimation of the Investment Factors

As explained in Section 4 and following Cunha et al. (2010), I consider skill investment choices made by the families to be sources of unobserved heterogeneity. In this Section, I describe the measures used for the identifications of the latent factors. In the Web Appendix, I present the estimation results.

I identify one investment factor per skill dimension. That is, I estimate a latent factor of investment in cognitive skills and another latent factor of investment in non-cognitive skills. To identify each investment factor, I need at least three manifest scores that relates to each investment dimension. Given that after hours tutoring is very popular in South Korea, I can use data on the cost and type of tutoring as manifest variables for the identification of the investment in cognitive skills factor. For the identification of the non-cognitive skills investment factor I use measures of good parenting collected in the KYP-JHSP. In the creation of the non-cognitive investment measures I used several variables and combined them in three indexes, namely parental abuse, parental control and parental harmony.

The *parental abuse* index is an aggregation of the answers to the following questions: i) I frequently see my parents verbally abuse each other; ii) I frequently see one of my parents beat the other one; iii) I am often verbally abused by parents; iv) I am often severely beaten by parents. The *parental control* index is created by aggregating the following: i) When I go out, my parents usually know where I am; ii) When I go out, my parents usually know whom I am with; iii) When I go out, my parents usually know what I do; iv) When I go out, my parents usually know when I return. Finally, the *parental harmony* index is created using the following questions: i) My parents and I try to spend much time together; ii) My parents always treat me with love and affection; iii) My parents and I understand each other well; iv) My parents and I candidly talk about everything; v) I frequently talk about my thoughts and what I experience away from home with my parents; vi) My parents and I have frequent conversations.

1 Dynamic Model Likelihood Function

The likelihood function described by the empirical strategy presented in Section 4.2.1 of the main paper is the following:

$$
\begin{aligned}
\mathcal{L} = \prod_{i=1}^{N} \int\int\int\int & \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta^{M}_{t+1}, \alpha^{\mathbf{M}}_{\mathbf{t+1}}\theta_{i\in c,t}, \alpha^{\nabla_{\theta_S}(d)}_{t+1}\nabla_{\psi,i\in c,t}(d)\right) \times \\
& \left[\begin{array}{l} f_{\vartheta^{1,M=1}_{t+1}}\left(\xi^{1}_{t+1} - \alpha^{T_{1,A}}_{t+1} g^{M=1}_{A,t+1}(\theta_t, I_t) - \alpha^{T_{1,B}}_{t+1} g^{M=1}_{B,t+1}(\theta_t, I_t)\right) \times \dots \\ \dots \times f_{\vartheta^{L,M=1}_{t+1}}\left(\xi^{L}_{t+1} - \alpha^{T_{LA}}_{t+1} g^{M=1}_{A,t+1}(\theta_t, I_t) - \alpha^{T_{LB}}_{t+1} g^{M=1}_{B,t+1}(\theta_t, I_t)\right) \end{array}\right]^{M} \times \\
& \left[1 - \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta^{M}_{t+1}, \alpha^{\mathbf{M}}_{\mathbf{t+1}}\theta_{i\in c,t}, \alpha^{\nabla_{\theta_S}(d)}_{t+1}\nabla_{\psi,i\in c,t}(d)\right)\right] \times \\
& \left[\begin{array}{l} f_{\vartheta^{1,M=0}_{t+1}}\left(\xi^{1}_{t+1} - \alpha^{T_{1,A}}_{t+1} g^{M=0}_{A,t+1}(\theta_t, I_t) - \alpha^{T_{1,B}}_{t+1} g^{M=0}_{B,t+1}(\theta_t, I_t)\right) \times \dots \\ \dots \times f_{\vartheta^{L,M=0}_{t+1}}\left(\xi^{L}_{t+1} - \alpha^{T_{LA}}_{t+1} g^{M=0}_{A,t+1}(\theta_t, I_t) - \alpha^{T_{LB}}_{t+1} g^{M=0}_{B,t+1}(\theta_t, I_t)\right) \end{array}\right]^{1-M} \times \\
& \times f_{\nu_{t+1}}\left(I_{A,i\in c,t+1} - \alpha^{A}_{A,t}\theta_{A,i\in c,t} - \alpha^{A}_{B,t}\theta_{B,i\in c,t}\right) \times f_{\nu_{t+1}}\left(I_{B,i\in c,t+1} - \alpha^{B}_{A,t}\theta_{A,i\in c,t} - \alpha^{B}_{B,t}\theta_{B,i\in c,t}\right) \\
& \times f_{e^{1}_{t}}\left(\mathbf{X}_{t,T_1}, T_{t,1}, \zeta^{A}, \zeta^{B}\right) \times \dots \times f_{e^{L}_{t}}\left(\mathbf{X}_{t,T_L}, T_{t,L}, \zeta^{A}, \zeta^{B}\right) \\
& \times \Delta F_{\theta^{A}_{t},\theta^{B}_{t}}\left(\zeta^{A}, \zeta^{B}\right) dF_{I_{A,t+1}}\left(\zeta^{I_A}\right) dF_{I_{B,t+1}}\left(\zeta^{I_B}\right)
\end{aligned}
$$

2 Results

2.1 Skills Identification

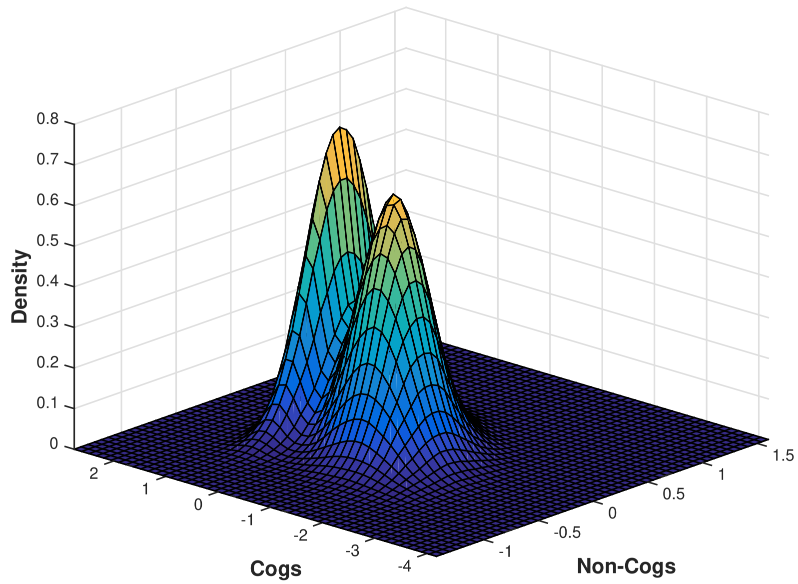
In this Section, I present the estimates of the first stage in which I identify the distribution of the cognitive and non-cognitive skills for t and $t+1$. Figures 1a and 1b show estimated distributions.

Table 1 presents the estimates for t in which I incorporate the structure proposed by Hansen et al. (2004) to address the problem of joint causality explained in Subsection 4.2.3 of the main paper. Based on the estimates in Table 1, in Figure 2, I decompose the test scores' variances into the fractions that are captured by observable characteristics, the latent factors and the residual respectively.

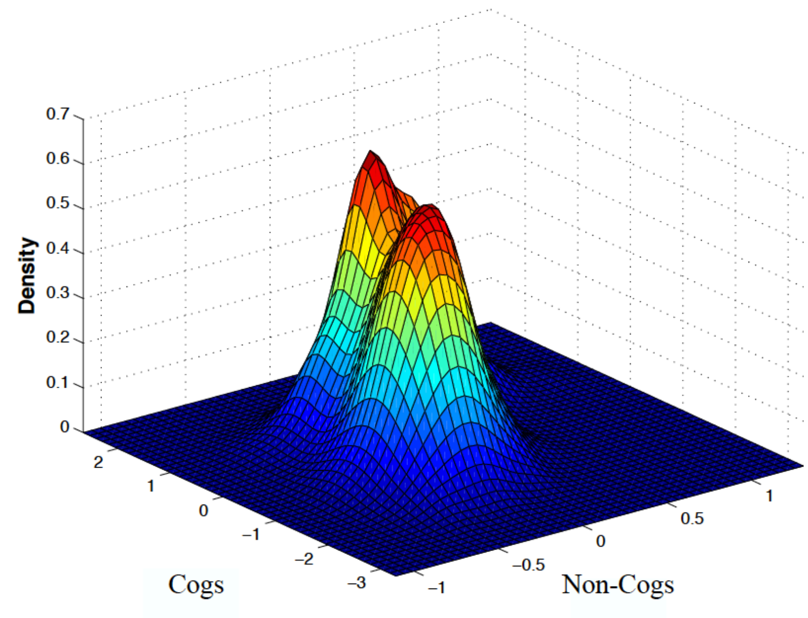
Figure 3 shows that although cognitive and non-cognitive skills are positively correlated, non-cognitive skills have a higher variance for the students that belong to the top deciles of the cognitive distribution. Therefore, among the relatively smart students, we can find a wider range of non-

Figure 1: Joint Skills Distribution

(a) $t = 1$



(b) $t = 2$



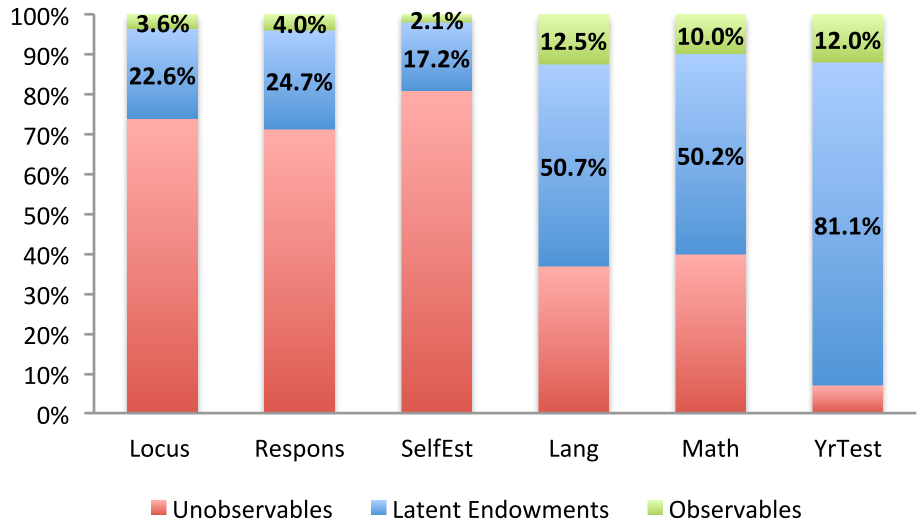
Note: Distributions obtained from parameters estimated in the structural model. In particular, estimating measurement systems (15) and (5) in the main paper. Complete set of estimates can be found in Tables 1 and 2. Factor distributions estimated using a mixture of two normals. The estimated parameters of those normals at $t = 1$ are: $\sigma_{1,NC} = 0.272$, $\sigma_{2,NC} = 0.630$, $\sigma_{1,C} = 0.321$, $\sigma_{2,C} = 0.217$, $\mu_{1,NC} - 0.076$, $\mu_{1,C} = -0.529$, $\rho_1 = 0.749$, $\rho_2 = 0.597$ and $p = 0.647$. For $t = 2$ are: $\sigma_{1,NC} = 0.235$, $\sigma_{2,NC} = 0.618$, $\sigma_{1,C} = 0.300$, $\sigma_{2,C} = 0.219$, $\mu_{1,NC} - 0.088$, $\mu_{1,C} = -0.570$, $\rho_1 = 0.797$, $\rho_2 = 0.610$ and $p = 0.613$, where p is the mixing probability.

Table 1: Identification of Skills at $t = 1$

		Locus of Control		Irresponsibility		Self Esteem		Language-SocSc.		Math-Sc.	Year Exam	
	Bullied	1	0	1	0	1	0	1	0	1/0	1	0
Age (Months)	0.010	-0.009	-0.010*	0.024**	0.013**	-0.006	-0.016***	-0.007	-0.009**	-0.009**	-0.017***	-0.012***
	(0.007)	(0.011)	(0.006)	(0.010)	(0.006)	(0.011)	(0.005)	(0.007)	(0.004)	(0.004)	(0.006)	(0.003)
Male	0.240***	0.201**	0.139***	-0.082	-0.057	0.319***	0.164***	-0.040	0.029	0.312***	0.010	-0.051**
	(0.051)	(0.079)	(0.040)	(0.076)	(0.040)	(0.078)	(0.039)	(0.054)	(0.028)	(0.030)	(0.043)	(0.022)
Older Siblings		0.018	0.018	0.027	-0.015	-0.027	-0.019	-0.111**	-0.026	0.031	0.019	-0.004
		(0.069)	(0.037)	(0.066)	(0.037)	(0.067)	(0.036)	(0.047)	(0.026)	(0.028)	(0.036)	(0.022)
Young Siblings		-0.060	0.035	-0.050	-0.070*	0.115	-0.016	0.047	0.084***	0.084***	0.044	0.085***
		(0.074)	(0.038)	(0.071)	(0.038)	(0.073)	(0.037)	(0.050)	(0.027)	(0.029)	(0.042)	(0.022)
lnInc_pc		0.068	0.075**	-0.139**	-0.086**	0.059	-0.000	0.090**	0.168***	0.143***	0.158***	0.112***
		(0.067)	(0.037)	(0.065)	(0.037)	(0.066)	(0.036)	(0.046)	(0.026)	(0.028)	(0.037)	(0.022)
Urban		0.053	0.209***	-0.077	-0.090	0.034	0.080	0.274***	0.057	0.068	0.031	-0.030
		(0.115)	(0.059)	(0.111)	(0.059)	(0.113)	(0.057)	(0.078)	(0.041)	(0.044)	(0.061)	(0.032)
Lives: 2 Parents		0.167	0.224*	-0.470**	-0.231*	0.298	0.202*	0.220*	0.364***	0.362***	0.249***	0.207***
		(0.191)	(0.118)	(0.183)	(0.118)	(0.186)	(0.116)	(0.129)	(0.083)	(0.089)	(0.096)	(0.076)
Lives: OnlyMom		0.584**	0.237	-0.583**	-0.138	0.853***	0.229	0.061	0.362***	0.339***	0.002	0.049
		(0.262)	(0.156)	(0.252)	(0.156)	(0.256)	(0.152)	(0.178)	(0.109)	(0.116)	(0.170)	(0.094)
Father Educ: 2yColl		0.278*	0.083	-0.120	-0.154*	0.199	-0.091	0.263**	0.120**	0.198***	0.227***	0.198***
		(0.155)	(0.079)	(0.148)	(0.079)	(0.151)	(0.077)	(0.105)	(0.055)	(0.059)	(0.084)	(0.044)
Father Educ: 4yColl		0.086	0.154***	-0.164*	-0.138***	0.025	0.105**	0.342***	0.312***	0.189***	0.189***	0.262***
		(0.093)	(0.047)	(0.090)	(0.047)	(0.091)	(0.046)	(0.063)	(0.033)	(0.036)	(0.051)	(0.027)
Father Educ: GS		0.073	0.328***	-0.059	-0.426***	0.058	0.207**	0.594***	0.432***	0.294***	0.247***	0.345***
		(0.153)	(0.085)	(0.147)	(0.085)	(0.149)	(0.082)	(0.104)	(0.059)	(0.061)	(0.075)	(0.044)
Non-Cognit.	-0.024	1.262***	1.151***	-1.220***	-1.196***	1.163***	1	0.760***	0.762***	0.848***		
	(0.190)	(0.143)	(0.090)	(0.145)	(0.096)	(0.170)	■	(0.136)	(0.099)	(0.112)		
Cognitive	-0.056							0.559***	0.547***	0.517***	1	1
	(0.053)							(0.042)	(0.026)	(0.027)	■	■
% Bullies	0.756***											
	(0.266)											
%Troub Fam	-4.177***											
	(1.333)											
%Troub Fam^2	5.474***											
	(1.654)											
Constant	-0.415	-0.531	-0.822***	1.079***	0.699***	-1.069***	-0.127	-0.797***	-1.244***	-1.257***	-0.939***	-0.690***
	(0.288)	(0.335)	(0.200)	(0.321)	(0.200)	(0.326)	(0.195)	(0.227)	(0.141)	(0.151)	(0.179)	(0.130)

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. *lnInc_pc* stands for log of household income per capita. *Lives: 2 Parents* takes the value of 1 if child lives with both parents and zero otherwise. *Lives: Only Mother* takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. *FatherEd* stands for father's education attainment. *FatherEd: GS* takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less. *% of Troubled Fam.* refers to the number of peers whose families score above the mean in the violent family index as described in footnote 18 of the main paper. *% of Bullies* refers to the number of peers that claim to have bullied a classmate. Factor distributions estimated using a mixture of two normals. . The estimated parameters of those normals are: $\sigma_{1,NC} = 0.272$, $\sigma_{2,NC} = 0.630$, $\sigma_{1,C} = 0.321$, $\sigma_{2,C} = 0.217$, $\mu_{1,NC} - 0.076$, $\mu_{1,C} = -0.529$, $\rho_1 = 0.749$, $\rho_2 = 0.597$ and $p = 0.647$, where p is the mixing probability. Observations: 3,097. Computational restrictions required the cognitive loadings on the *Year Exam* score not to differ across victimization conditions and normalized to 1. They also required the *Math-Science* score not to have vicitmization-specific parameters.

Figure 2: Decomposing Variances of Test Scores at $t = 1$



cognitive traits that could potentially yield students that are cognitively and non-cognitively well endowed and students that are smart and and have a difficult time building social relations.

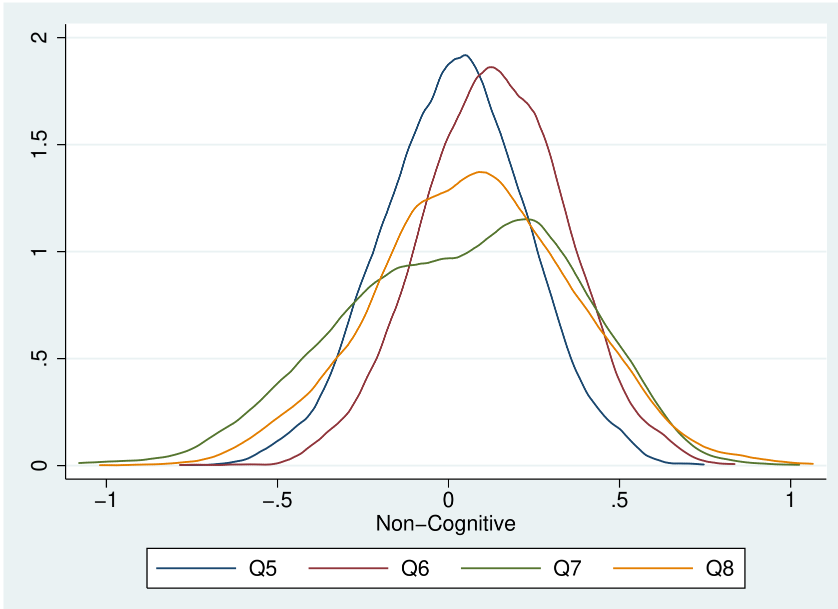
Finally, Table 2 presents the complete set of estimates for the test scores in $t + 1$.

Table 2: Identification of Skills at $t = 2$

VARIABLES	Locus	Irrespons	Self-est	Lang-SSc	Math-Scie	YearExam
Age (months)	-0.023***	0.008	-0.008	-0.013***	-0.008*	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Male	0.112***	-0.067*	0.136***	0.059*	0.366***	-0.058**
	(0.038)	(0.038)	(0.038)	(0.033)	(0.033)	(0.026)
Older Siblings	0.034	-0.016	0.055	-0.004	0.013	0.023
	(0.036)	(0.036)	(0.036)	(0.032)	(0.032)	(0.027)
Young Siblings	0.041	-0.092**	0.078**	0.143***	0.125***	0.082***
	(0.037)	(0.037)	(0.037)	(0.033)	(0.033)	(0.029)
lnInc_pc	0.087**	-0.039	0.068*	0.158***	0.164***	0.171***
	(0.037)	(0.037)	(0.037)	(0.034)	(0.033)	(0.030)
Urban	0.099*	-0.005	0.023	0.084*	0.059	-0.099**
	(0.058)	(0.058)	(0.058)	(0.050)	(0.050)	(0.039)
Lives: Both Parents	-0.079	-0.186**	0.062	0.286***	0.411***	0.302***
	(0.087)	(0.087)	(0.088)	(0.076)	(0.075)	(0.062)
Lives: Only Mother	0.022	-0.244*	0.141	0.068	0.214*	0.220**
	(0.132)	(0.132)	(0.132)	(0.114)	(0.113)	(0.092)
Father Edu: 2yColl	-0.004	-0.209***	0.102	0.088	0.187***	0.178***
	(0.075)	(0.075)	(0.075)	(0.066)	(0.065)	(0.052)
Father Edu: 4yColl	0.112**	-0.166***	0.105**	0.295***	0.219***	0.253***
	(0.045)	(0.045)	(0.045)	(0.039)	(0.039)	(0.030)
Father Edu: GS	0.211**	-0.245***	0.119	0.358***	0.304***	0.346***
	(0.086)	(0.086)	(0.086)	(0.076)	(0.075)	(0.058)
Non-Cogn. Factor	1.190***	-1.325***	1	1.351***	1.160***	
	(0.109)	(0.131)	■	(0.215)	(0.179)	
Cognitive Factor				0.405***	0.461***	1
				(0.039)	(0.034)	■
Constant	-0.328	0.445**	-0.499**	-1.125***	-1.419***	-0.937***
	(0.200)	(0.201)	(0.201)	(0.181)	(0.180)	(0.163)

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. *lnInc_pc* stands for log of household income per capita. *Lives: Both Parents* takes the value of 1 if child lives with both parents and zero otherwise. *Lives: Only Mother* takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. *FatherEd* stands for father's education attainment. *FatherEd: GS* takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less. Factor distributions estimated using a mixture of two normals. The estimated parameters of those normals are: $\sigma_{1,NC} = 0.235$, $\sigma_{2,NC} = 0.618$, $\sigma_{1,C} = 0.300$, $\sigma_{2,C} = 0.219$, $\mu_{1,NC} - 0.088$, $\mu_{1,C} = -0.570$, $\rho_1 = 0.797$, $\rho_2 = 0.610$ and $p = 0.613$, where p is the mixing probability. Observations: 2,731.

Figure 3: Distribution of Non-cognitive by Decile of Cognitive Skills at $t = 1$



Note: Non-cognitive skills kernel densities for selected deciles of cognitive skills.

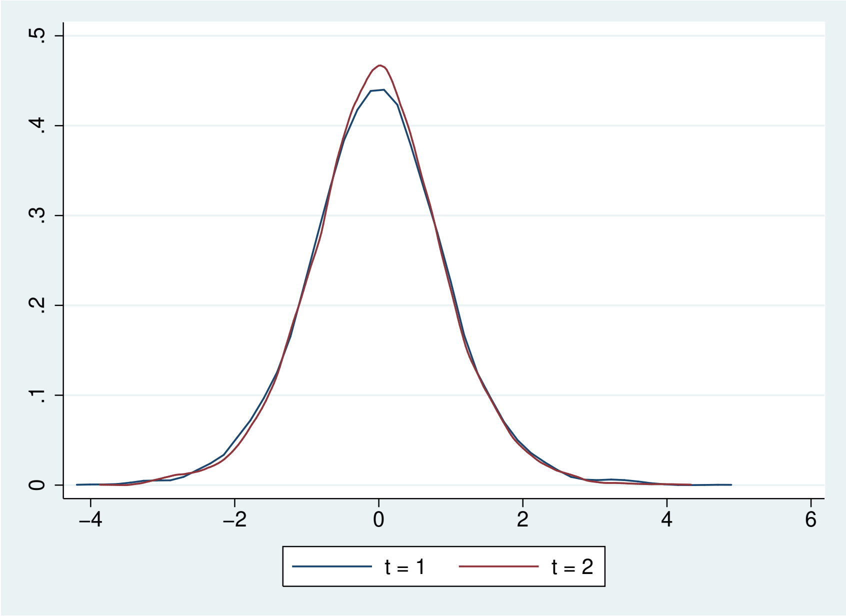
2.2 Estimation of Investment Factors

I estimate the measurement system described by equations (5) and (6) in the main paper and obtain the underlying distributions from which the unobserved heterogeneity in investment comes from. Figure 4a show that investment in non-cognitive skills is remarkably stable across waves. Figure 4b shows two important characteristics of investment in cognitive skills. First, its bimodality. That may be the case because there are a proportion of kids that take no tutoring at all. Second, investment in cognitive skills is not stable in time. This responds to the fact that participation private tutoring falls as kids grow up.

Tables 3 and 4 show that the non-cognitive investment factor closely relates with good parental practices as it correlates positively with parental control and negatively with physical and verbal abuse. In the same way, the cognitive investment factor relates with the quality of after-class tutoring. It is positively correlated with how private the tutoring is, and how many hours the student spends in such after-class activities.

Figure 4: Unobserved Investment Factors

(a) Non-Cognitive Investment Factor



(b) Cognitive Investment Factor

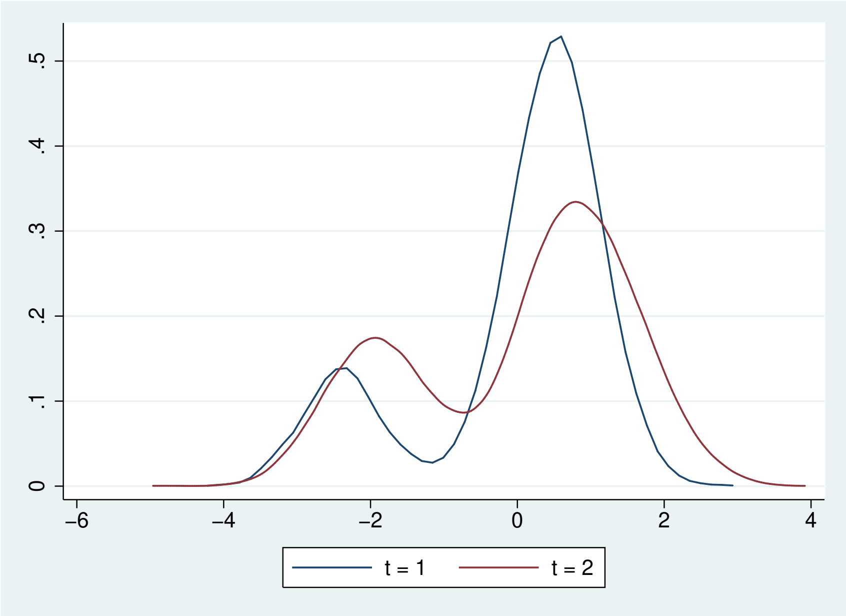


Table 3: Identification of Unobserved Non-Cognitive Investment Factor

	$t=1$			$t=2$		
VARIABLES	Abuse	Control	Harmony	Abuse	Control	Harmony
Age (months)	0.0001	0.0005	-0.0081**	-0.0024	-0.0015	-0.0036
	(0.002)	(0.005)	(0.004)	(0.002)	(0.005)	(0.004)
Male	0.0434***	-0.2661***	-0.1638***	0.0169	-0.3143***	-0.2137***
	(0.016)	(0.033)	(0.026)	(0.014)	(0.034)	(0.028)
Older Siblings	-0.0019	-0.0153	0.0212	-0.0131	-0.0632*	-0.0025
	(0.015)	(0.031)	(0.022)	(0.013)	(0.033)	(0.028)
Young Siblings	-0.0079	0.0335	0.0186	-0.0017	-0.0136	0.0008
	(0.015)	(0.032)	(0.025)	(0.014)	(0.033)	(0.027)
lnInc_pc	-0.0217	0.0474	0.0899***	-0.0316**	0.1254***	0.1030***
	(0.015)	(0.033)	(0.026)	(0.014)	(0.033)	(0.027)
Urban	-0.0151	0.0256	0.0778**	-0.0277	0.1167**	0.1202***
	(0.024)	(0.050)	(0.037)	(0.021)	(0.051)	(0.044)
Lives: Both Parents	-0.1225***	0.1385*	0.1136*	-0.1082***	0.1474*	0.2351***
	(0.036)	(0.076)	(0.062)	(0.032)	(0.079)	(0.065)
Lives: Only Mother	-0.1391**	0.1400	0.2128**	-0.0861*	0.0581	0.3037***
	(0.054)	(0.113)	(0.086)	(0.045)	(0.109)	(0.087)
Father Edu: 2yColl	0.0555*	0.0407	0.1361***	0.0346	-0.0107	-0.0066
	(0.031)	(0.065)	(0.047)	(0.027)	(0.065)	(0.051)
Father Edu: 4yColl	-0.0287	0.0934**	0.0614**	-0.0411**	0.1372***	0.0998***
	(0.019)	(0.039)	(0.030)	(0.016)	(0.040)	(0.032)
Father Edu: GS	-0.1132***	0.3694***	0.1430**	-0.0750**	0.2273***	0.0854
	(0.034)	(0.072)	(0.059)	(0.030)	(0.072)	(0.055)
Non-Cogn Invest.	-0.1268***	0.5843***	1	-0.1269***	0.5564***	1
	(0.009)	(0.017)	.	(0.008)	(0.018)	.
Constant	2.1087***	-0.2924*	-0.4708***	2.1324***	-0.6303***	-0.6766***
	(0.082)	(0.175)	(0.147)	(0.072)	(0.177)	(0.152)
Observations	2,988			2,968		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. *lnInc_pc* stands for log of household income per capita. *Lives: Both Parents* takes the value of 1 if child lives with both parents and zero otherwise. *Lives: Only Mother* takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. *FatherEd* stands for father's education attainment. *FatherEd: GS* takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less. Factor distributions estimated using a mixture of two normals.

Table 4: Identification of Unobserved Cognitive Investment Factor

	$t=1$			$t=2$		
VARIABLES	Type Tutor	Tutor Time	Exp Tutor	Type Tutor	Tutor Time	Exp Tutor
Age (months)	0.0044	-0.0090*	0.0020	0.0056	-0.0101**	-0.0036
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Male	-0.0204	-0.0320	-0.0382	0.0999***	0.0830**	0.0962**
	(0.031)	(0.035)	(0.025)	(0.033)	(0.035)	(0.041)
Older Siblings	-0.0548*	-0.0016	-0.0329	-0.0619*	-0.0218	-0.0032
	(0.030)	(0.033)	(0.026)	(0.032)	(0.034)	(0.040)
Young Siblings	0.0043	0.0570*	0.0756***	0.0147	0.0502	0.1424***
	(0.030)	(0.034)	(0.024)	(0.033)	(0.035)	(0.042)
lnInc_pc	0.1166***	0.1539***	0.2429***	0.1102***	0.1233***	0.3401***
	(0.032)	(0.035)	(0.027)	(0.036)	(0.037)	(0.051)
Urban	-0.0846*	-0.1092**	-0.2407***	-0.2417***	-0.2372***	-0.5036***
	(0.048)	(0.053)	(0.039)	(0.051)	(0.054)	(0.062)
Lives: Both Parents	0.1304	-0.0170	-0.0108	0.2619***	0.2068**	0.2958**
	(0.094)	(0.103)	(0.085)	(0.095)	(0.098)	(0.134)
Lives: Only Mother	0.1159	-0.0154	0.0512	0.1271	0.1678	0.3272*
	(0.121)	(0.133)	(0.104)	(0.121)	(0.126)	(0.167)
Father Edu: 2yColl	-0.0247	0.1701**	0.1167**	-0.0264	0.0168	-0.0446
	(0.062)	(0.069)	(0.049)	(0.064)	(0.068)	(0.078)
Father Edu: 4yColl	0.0485	0.1247***	0.0477	0.0335	0.0912**	-0.0260
	(0.037)	(0.041)	(0.030)	(0.039)	(0.042)	(0.048)
Father Edu: GS	-0.0867	0.1108	0.0100	0.1455**	0.2851***	0.1395
	(0.067)	(0.075)	(0.051)	(0.073)	(0.076)	(0.096)
Cogn Investment	0.4747***	0.3025***	1	0.4185***	0.3387***	1
	(0.013)	(0.014)	.	(0.011)	(0.012)	.
Constant	-0.5612***	-0.5452***	-0.7926***	-0.5842***	-0.5216***	-1.2945***
	(0.176)	(0.191)	(0.167)	(0.196)	(0.201)	(0.297)
Observations	2,918			2,761		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. *lnInc_pc* stands for log of household income per capita. *Lives: Both Parents* takes the value of 1 if child lives with both parents and zero otherwise. *Lives: Only Mother* takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. *FatherEd* stands for father's education attainment. *FatherEd: GS* takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less. Factor distributions estimated using a mixture of two normals.

2.3 Incidence of victimization

Table 5 presents the coefficient estimates of the likelihood of being bullied from which the marginal effects presented in Table 6 were calculated.

Table 5: Likelihood of Being Bullied

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age (months)	0.005	0.006	0.006	0.005	0.006	0.006	0.006
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Male	0.268***	0.267***	0.265***	0.254***	0.253***	0.251***	0.247***
	(0.065)	(0.065)	(0.065)	(0.066)	(0.066)	(0.066)	(0.066)
Young Siblings	-0.093*	-0.091*	-0.091*	-0.091*	-0.090*	-0.090*	-0.090*
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
% Trouble Fams.	-2.405	-2.533	-2.326	-2.513	-2.655	-2.448	-2.003
	(1.786)	(1.791)	(1.797)	(1.784)	(1.789)	(1.796)	(1.837)
% Trouble Fams.^{2}	3.336	3.473*	3.230	3.369*	3.516*	3.279	2.859
	(2.052)	(2.057)	(2.064)	(2.046)	(2.052)	(2.059)	(2.092)
% Bullies				0.786	0.820	0.784	0.699
				(0.543)	(0.545)	(0.546)	(0.550)
Mass[selfest]	-0.040**	-0.043***	-0.041**	-0.039**	-0.043***	-0.041**	-0.039***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Mass[ClassScore]		0.022	0.028		0.024	0.029	0.036
		(0.022)	(0.022)		(0.022)	(0.022)	(0.023)
Mass[Income]			-0.021*			-0.020	-0.018
			(0.012)			(0.012)	(0.012)
Class Size							-0.010
							(0.009)
Non-Cogn	-0.387**	-0.388**	-0.385**	-0.381**	-0.382**	-0.379**	-0.378**
	(0.157)	(0.157)	(0.158)	(0.157)	(0.157)	(0.157)	(0.157)
Cognitive	0.045	0.042	0.041	0.044	0.041	0.039	0.037
	(0.047)	(0.048)	(0.048)	(0.047)	(0.048)	(0.048)	(0.048)
Constant	-0.853**	-0.874**	-0.856**	-0.876**	-0.900**	-0.881**	-0.672
	(0.385)	(0.386)	(0.387)	(0.385)	(0.386)	(0.387)	(0.425)
Joint Significance of Instruments							
χ^2	4.74	5.00	4.46	6.28	6.69	6.02	5.55
Pr > χ^2	0.0295	0.0254	0.0348	0.0122	0.0097	0.0141	0.0185
Log-Likelihood	-24833.98	-24833.47	-24832.04	-24829.39	-24828.79	-24827.46	-24826.76
Observations	2,874	2,874	2,874	2,874	2,874	2,874	2,874

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates equation (15) in the structural model. Estimations include region fixed-effects and observable controls age, gender and family composition. Mass[] refers to the number of observations within a window of 10% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. % of Troubled Families refers to the number of peers whose families score above the mean in the violent family index as described in footnote 19. % of Bullies refers to the number of peers that claim to have bullied a classmate.

Table 6: Likelihood of Being Bullied (Marginal Effects at the Mean)

	(1)	(2)	(3)	(4)	(5)	(6)
Non-Cogs	-0.070**	-0.071**	-0.070**	-0.069**	-0.069**	-0.069**
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Cognitive	0.008	0.008	0.007	0.008	0.007	0.007
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Mass[selfest]	-0.007**	-0.008***	-0.007**	-0.007**	-0.008***	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mass[ClassScore]		0.004	0.005		0.004	0.005
		(0.004)	(0.004)		(0.004)	(0.004)
Mass[Income]			-0.004*			-0.004
			(0.002)			(0.002)
Additional Controls						
% Trouble Fams.	X	X	X	X	X	X
% Bullies				X	X	X

Note: *** p<0.01, ** p<0.05, * p<0.1. Estimates equation (15) of the structural model. See complete estimates in Table 5. Mass[] refers to the number of observations within a window of 10% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. % of Troubled Families refers to the number of peers whose families score above the mean in the violent family index as described in footnote 18 of the main paper. % of Bullies refers to the number of peers that claim to have bullied a classmate.

2.4 Effects of Bullying on Skill Production: Understanding the size of the effect on $\theta_{NC,t+1}$

In this Appendix I present some results that help understand the impacts found in the paper using understandable metrics. I estimate the following specification:

$$Y = \mathbf{X}_Y \beta^Y + \alpha^{Y,NC}\theta^{NC} + \alpha^{Y,C}\theta^{C} + e^Y$$

Its purpose is to capture the effect of skills on more tangible outcomes, and in that way have a better picture about how the skills lost to bullying hurt the development of successful lives. See Sarzosa and Urzua (2021) for a detailed explanation on how the outcome measures were constructed.

Table 7: Effect of unobserved heterogeneity at age 16

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Probit					LS		
VARIABLES	College	LifeSatisf	Healthy	Smoke	Drink	Depression	StressImage	StressFiends	StressSchool	StressTotal
Age (months)	0.0040	0.0139	0.0079	-0.0002	-0.0125*	-0.0009	0.0045	0.0054	-0.0023	0.0023
	(0.008)	(0.011)	(0.007)	(0.010)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Male	-0.3658***	0.3862***	0.0930*	0.4307***	-0.0347	-0.2526***	-0.3412***	0.0875**	-0.2591***	-0.1685***
	(0.057)	(0.093)	(0.054)	(0.074)	(0.052)	(0.037)	(0.037)	(0.039)	(0.037)	(0.037)
Older Siblings	-0.0005	0.1973**	0.0333	-0.0918	-0.0197	-0.0094	0.0224	0.0463	0.0641*	0.0408
	(0.054)	(0.082)	(0.052)	(0.068)	(0.051)	(0.036)	(0.036)	(0.038)	(0.036)	(0.036)
Young Siblings	0.0405	0.1529*	-0.0233	-0.1841**	-0.0967*	-0.0013	-0.0172	0.0492	0.0726**	0.0284
	(0.055)	(0.082)	(0.053)	(0.073)	(0.051)	(0.036)	(0.037)	(0.038)	(0.037)	(0.037)
lnInc_pc	0.0811	0.3043***	0.1302**	0.0366	-0.1029**	0.0073	-0.0139	0.0330	0.0574	0.0319
	(0.056)	(0.088)	(0.053)	(0.067)	(0.052)	(0.037)	(0.037)	(0.038)	(0.037)	(0.037)
Urban	-0.1443*	0.0536	-0.1025	0.1205	-0.0566	-0.0354	-0.0007	-0.1488***	0.1336**	0.0273
	(0.086)	(0.118)	(0.079)	(0.108)	(0.077)	(0.054)	(0.055)	(0.057)	(0.055)	(0.055)
Lives: Both Parents	0.3620***	0.0954	0.3392***	-0.3425**	-0.3686***	-0.0692	-0.1907**	-0.1186	0.1786**	-0.0420
	(0.132)	(0.190)	(0.130)	(0.147)	(0.120)	(0.088)	(0.088)	(0.091)	(0.088)	(0.088)
Lives: Only Mother	0.2993	0.3979	0.2900	0.1326	0.0274	-0.0950	-0.3407***	-0.2475*	-0.0591	-0.3171***
	(0.185)	(0.269)	(0.180)	(0.196)	(0.168)	(0.121)	(0.123)	(0.127)	(0.122)	(0.123)
Father Edu: 2yColl	0.1782	0.2736*	-0.2132**	-0.4097**	-0.1278	0.0093	0.0032	0.0091	0.1268*	0.0910
	(0.119)	(0.161)	(0.104)	(0.164)	(0.103)	(0.072)	(0.072)	(0.075)	(0.072)	(0.073)
Father Edu: 4yColl	-0.0196	0.2157**	-0.0199	-0.1406*	-0.0358	0.0272	-0.0990**	-0.0505	0.0910**	0.0013
	(0.068)	(0.099)	(0.064)	(0.084)	(0.062)	(0.044)	(0.044)	(0.046)	(0.044)	(0.044)
Father Edu: GS	-0.1255	0.5521***	0.0177	-0.4167**	-0.1610	-0.0339	-0.2672***	-0.1435	0.0037	-0.1677**
	(0.125)	(0.200)	(0.120)	(0.187)	(0.122)	(0.084)	(0.085)	(0.087)	(0.084)	(0.085)
ParentWantsColl	0.6257***	0.1989	0.0858	-0.2264**	-0.0391	0.0237	0.0287	-0.0082	0.4518***	0.1969***
	(0.090)	(0.136)	(0.090)	(0.107)	(0.087)	(0.061)	(0.062)	(0.064)	(0.062)	(0.062)
ParentWantsGS	0.5497***	0.2426	0.0218	-0.2332	-0.0226	0.1340*	0.0805	0.0956	0.5739***	0.3283***
	(0.115)	(0.172)	(0.112)	(0.143)	(0.109)	(0.077)	(0.078)	(0.081)	(0.078)	(0.078)
Non-Cogs	-0.0726	3.5356***	1.0919***	-0.6746***	-0.5552***	-1.9588***	-1.5315***	-1.2128***	-1.3627***	-1.7103***
	(0.169)	(0.676)	(0.190)	(0.231)	(0.163)	(0.111)	(0.110)	(0.114)	(0.104)	(0.106)
Cognitive	0.0893**	-0.2060***	-0.1937***	-0.1479***	-0.0570	0.2313***	0.1870***	0.1199***	0.1079***	0.1391***
	(0.041)	(0.077)	(0.043)	(0.045)	(0.037)	(0.026)	(0.027)	(0.028)	(0.026)	(0.026)
Constant	-0.4744	-2.1872***	-1.0961***	-1.0815***	0.7336***	0.1578	0.3884*	-0.0374	-0.9150***	-0.2890
	(0.296)	(0.502)	(0.291)	(0.357)	(0.276)	(0.196)	(0.199)	(0.205)	(0.197)	(0.198)
Observations	2,345	2,685	2,685	2,685	2,685	2,685	2,676	2,678	2,678	2,654

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

lnInc_pc stands for log of household income per capita. *Lives: Both Parents* takes the value of 1 if child lives with both parents and zero otherwise. *Lives: Only Mother* takes the value of 1 if child lives only with her motherand zero otherwise. The excluded category is living only with the father or living with no parent. *FatherEd* stands for father's education attainment. *FatherEd: GS* takes the value of 1 if the father holds a graduate degree and zero otherwise. The excluded category is fathers with high school or less.

3 Robustness

3.1 Testing for Threats to Identification: Parental Shocks

Identification of the potential outcomes model and its associated treatment parameters requires $e^M_{t+1} \perp \left(\eta^0_{A,t+1}, \eta^0_{B,t+1}, \eta^1_{A,t+1}, \eta^1_{B,t+1}\right)$. Thus, shocks that simultaneously alter the 'sorting into victimization' and period $t+1$ skills without going through $g^M_{S,t+1}(\theta_t, I_t)$ are considered threats to the identification of the model. In this Appendix, I deal with this concern. I do so by using information on parental shocks reported in the survey. In wave three (model period $t+2$), students are asked about numerous past life events. Fortunately, they are also asked to locate those event in time. That way, I was able to collect shocks that happened only between times t and $t+1$. The life events I considered were whether any of the student's parents had died, whether either parent had failed in business or lost a job, and whether either parent had been hospitalized. All these shocks have the potential of increasing the cances of being victimized, and affect skill accumulation between t and $t+1$.

Using these data on parental shocks, I estimate a version of the model that includes them in the victimization equation and in the potential outcome equations as an additional observable control. Controlling for those shocks will arguably make the identifying assumption less likely to be violated. Thus, if the production functions' estimated parameters in the model with shocks differ significantly relative to those in the model without parental shocks, then it would be evidence of the latter model—the one used in the paper—being misspecified and yielding biased results.

Table (8) in this Appendix shows that having suffered a parental shock significantly increases the chances of being victimized. Students who lost their parents, whose parents had an economic loss, or whose parents became seriously ill are 9.7 percentage points more likely to be bullied. That is, a students that suffered those shocks is almost twice more likely to be bullied than the average student. However, despite that significant relation between parental shocks and victimization, Table (8) shows that they do not affect the estimation of the skill production functions. The results are very similar to the ones presented in the main model. This fact indicates that the setting and empirical strategy employed in the paper are such that parental shocks' scope as drivers of skills accumulation is too small to be meaningful.

Table 8: Dynamic Estimation

	(1)	(2)	(3)		(4)	(5)	(6)	(7)
					$M_{t+1}=0$		$M_{t+1}=1$	
	M_{t+1}	$I_{NC,t+1}$	$I_{C,t+1}$		$\theta_{NC,t+1}$	$\theta_{C,t+1}$	$\theta_{NC,t+1}$	$\theta_{C,t+1}$
$\theta_{NC,t}$	-0.069**	0.677***	0.107	$\theta_{NC,t}$	0.953	0.106	0.907	0.091
	(0.028)	(0.075)	(0.110)		(0.016)	(0.016)	(0.040)	(0.041)
$\theta_{C,t}$	0.008	0.027	0.342***	I_{t+1}	0.030	0.057	0.030	0.048
	(0.009)	(0.022)	(0.036)		(0.012)	(0.008)	(0.028)	(0.023)
$\nabla\left(\widehat{\theta}_{NC,t}\right)$	-0.009**			ρ	-0.084	-0.036	0.401	-0.171
	(0.003)				(0.086)	(0.042)	(0.359)	(0.143)
$\nabla\left(\widehat{\theta}_{C,t}\right)$	0.005							
	(0.004)							
$\nabla\left(\text{Inc}_t\right)$	-0.004							
	(0.002)							
Shocks	0.097**							
	(0.038)							

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) presents the marginal effects of the estimation of equation $M_{t+1} = \mathbf{1}\left[\mathbf{X}_{t+1,M}\beta^M_{t+1} + \Lambda^M_{t+1}\Theta'_{i\in c,t} + \Lambda^{M_c}_{t+1}\nabla_{\psi,i\in c}(d) + \Gamma Z_{t+1,c} > e^M_{t+1}\right]$ of the structural model. It includes observable controls age, gender, family composition and % of troubled families (i.e., the number of peers whose families score above the mean in the violent family index). $\nabla(\cdot)$ refers to the number of classmates within a window of 10% of a SD around observation i. $\widehat{\theta}_{NC}$ is the residualized measure of self-esteem and $\widehat{\theta}_C$ is the residualized measure of the yearly test. The marginal effect of the $\nabla(\cdot)$ variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. Columns (2) and (3) preset the estimates of equations $I_{S,i\in c,t+1} = \alpha^S_{A,t}\theta_{A,i\in c,t} + \alpha^S_{B,t}\theta_{B,i\in c,t} + \varepsilon_{S,i\in c,t+1}$ for $S \in \{A,B\}$ of the structural model. Columns (4)-(7) present the estimates of the victimization-specific production functions of non-cognitive and cognitive skills

$$\theta_{S,i\in c,t+1} = \begin{cases} \left[\gamma^0_{A,S,t}\theta^{\rho^0_S}_{A,i\in c,t} + \gamma^0_{B,S,t}\theta^{\rho^0_S}_{B,i\in c,t} + \gamma^0_{I,S,t}I^{\rho^0_S}_{S,i\in c,t+1}\right]^{1/\rho^0_S} + \eta^0_{S,i\in c,t} & \text{if } M_{i,t+1} = 0 \\ \left[\gamma^1_{A,S,t}\theta^{\rho^1_S}_{A,i\in c,t} + \gamma^1_{B,S,t}\theta^{\rho^1_S}_{B,i\in c,t} + \gamma^1_{I,S,t}I^{\rho^1_S}_{S,i\in c,t+1}\right]^{1/\rho^1_S} + \eta^1_{S,i\in c,t} & \text{if } M_{i,t+1} = 1 \end{cases}$$

for $S \in \{A,B\}$. Note that the coefficient for $\theta_{C,t}$ (i.e., $\gamma_{C,t}$) can be obtained from $\gamma_{C,t} = 1 - \gamma_{NC,t} - \gamma_{I,t}$.

3.2 Treating Cognitive Skills as Impervious to Victimization

Bullying *might* not affect the accumulation of cognitive skills because cognitive skills are *less malleable* than non-cognitive skills during adolescence (Walsh, 2004; Kautz et al., 2014). In this Appedix, I present the estimates of a version of the model in which cognitive skills are allowed to evolve, but are not subject to the effects of victimization. That is

$$\theta_{A,i\in c,t+1} = \begin{cases} \left[\gamma^0_{A,A,t}\theta^{\rho^0_A}_{A,i\in c,t} + \gamma^0_{B,A,t}\theta^{\rho^0_A}_{B,i\in c,t} + \gamma^0_{I,A,t}I^{\rho^0_A}_{A,i\in c,t+1}\right]^{1/\rho^0_A} + \eta^0_{A,i\in c,t} & \text{if } M_{i,t+1} = 0 \\ \left[\gamma^1_{A,A,t}\theta^{\rho^1_A}_{A,i\in c,t} + \gamma^1_{B,A,t}\theta^{\rho^1_A}_{B,i\in c,t} + \gamma^1_{I,A,t}I^{\rho^1_A}_{A,i\in c,t+1}\right]^{1/\rho^1_A} + \eta^1_{A,i\in c,t} & \text{if } M_{i,t+1} = 1 \end{cases} \tag{1}$$

$$\theta_{B,i\in c,t+1} = \left[\gamma_{A,B,t}\theta^{\rho_B}_{A,i\in c,t} + \gamma_{B,B,t}\theta^{\rho_B}_{B,i\in c,t} + \gamma_{I,B,t}I^{\rho_B}_{B,i\in c,t+1}\right]^{1/\rho_B} + \eta_{B,i\in c,t} \tag{2}$$

$$I_{S,i\in c,t+1} = \alpha^S_{A,t}\theta_{A,i\in c,t} + \alpha^S_{B,t}\theta_{B,i\in c,t} + \varepsilon_{S,i\in c,t+1} \qquad \text{for } S \in \{A, B\} \tag{3}$$

$$M_{i\in c,t+1} = \mathbf{1}\left[\mathbf{X}_{it}\beta^M_{t+1} + \alpha^{M_A}_{t+1}\theta_{A,i\in c,t} + \alpha^{M_B}_{t+1}\theta_{B,i\in c,t} + \Lambda^{M_c}_{t+1}\nabla_{\psi,i\in c}(d) + \Gamma Z_c > e^M_{i\in c,t+1}\right] \tag{4}$$

where the notation follows that of the main model in the paper. That is, $\gamma^{M_i}_{I,S,t} = 1 - \gamma^{M_i}_{A,S,t} - \gamma^{M_i}_{B,S,t}$ for the victimization status $M_i \in \{0, 1\}$, $\mathbf{1}[\cdot]$ is an indicator function that takes the value of 1 if true . $\eta^{M_i}_{A,i\in c,t}$ denote shocks that affect the accumulation of non-cognitive skills by victimization status between t and $t+1$, and $\eta_{B,i\in c,t}$ denotes the shocks that affect the accumulation of cognitive skills between t and $t+1$. The CES parameters contain a superscript $M_i \in \{0, 1\}$ to indicate that the skills production functions of non-cognitive skills for victimized students are different from those of non-victimized ones.

The results do not differ greatly from the main model in the paper. Table 10 shows the overall average effect of victimization on non-cognitive skills is -0.203 (versus -0.249 in the main model), and the effect on the average student is -0.206 (versus -0.257 in the main model). When, I estimate the ATE for each level of skills, Figure 5 shows a similar pattern to the results of the main model (Figure 6a in the paper), where the students that start the process with low stocks of skills face harsher consequences of victimization in the production of period t+1 non-cognitive skills. The constrained model yields an ATE with a slightly higher gradient on the cognitive dimension. That is, in the constrained model students with high initial cognitive skills have very small ATEs (around -0.06), while in the main model these students faced ATEs in the order of -0.17.

Table 9: Dynamic Estimation of Contrained Model

	(1)	(2)	(3)		(4)	(5)	(6)
					$M_{t+1}=0$	$M_{t+1}=1$	
	M_{t+1}	$I_{NC,t+1}$	$I_{C,t+1}$		$\theta_{NC,t+1}$	$\theta_{NC,t+1}$	$\theta_{C,t+1}$
$\theta_{NC,t}$	-0.071**	0.684***	0.113	$\theta_{NC,t}$	0.947	0.945	0.125
	(0.028)	(0.076)	(0.110)		(0.023)	(0.018)	(0.015)
$\theta_{C,t}$	0.008	0.030	0.340***	I_{t+1}	0.097	-0.011	0.056
	(0.009)	(0.023)	(0.036)		(0.016)	(0.013)	(0.008)
$\nabla\left(\widehat{\theta}_{NC,t}\right)$	-0.007***			ρ	0.089	0.009	-0.049
	(0.003)				(0.106)	(0.097)	(0.039)
$\nabla\left(\widehat{\theta}_{C,t}\right)$	0.005						
	(0.004)						
$\nabla\left(\mathrm{Inc}_t\right)$	-0.004*						
	(0.002)						

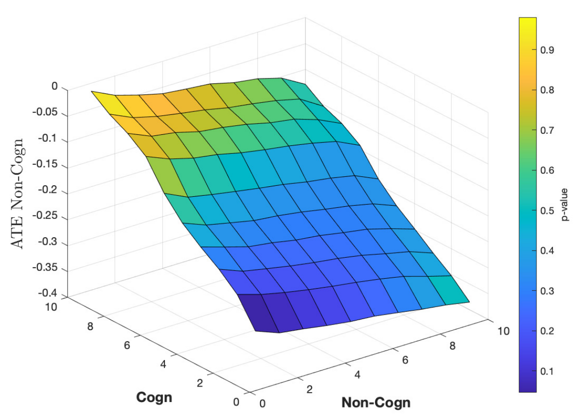
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) presents the marginal effects of the estimation of equation (4). It includes observable controls age, gender, family composition and % of troubled families (i.e., the number of peers whose families score above the mean in the violent family index). $\nabla(\cdot)$ refers to the number of classmates within a window of 10% of a SD around observation i. $\widehat{\theta}_{NC}$ is the residualized measure of self-esteem and $\widehat{\theta}_C$ is the residualized measure of the yearly test. The marginal effect of the $\nabla(\cdot)$ variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. Columns (2) and (3) preset the estimates of equations (3) for $S \in \{A, B\}$. Columns (4) and (5) present the estimates of the victimization-specific production functions of non-cognitive skills (1). And Column (6) presents the estimates of the production funtion of cognitive skills (2). Note that the coefficient for $\theta_{C,t}$ (i.e., $\gamma_{C,t}$) can be obtained from $\gamma_{C,t} = 1 - \gamma_{NC,t} - \gamma_{I,t}$.

Table 10: ATE of Being Bullied on Next Period Skills

	$\theta_{NC,t+1}$	
	$\mathbb{E}_{\theta_t}\left[ATE(\theta_t)\right]$	$ATE(\theta_t=\bar{\theta}_t)$
Estimated	-0.249***	-0.257***
	(0.020)	(0.020)
As SD of $\theta_{S,t+1}$	-0.399	-0.413

Note: Standard errors in parentheses. $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$. Let $ATE(\theta_t) = E\left[\widehat{\theta}^S_{t+1}|\theta_t^{NC},\theta_t^N,M_{t+1}=1\right] - E\left[\widehat{\theta}^S_{t+1}|\theta_t^{NC},\theta_t^N,M_{t+1}=0\right]$ for $S\in\{NC,C\}$. The Table present the mean average treatment affect $\mathbb{E}_{\theta_t}\left[ATE(\theta_t)\right]=\int ATE(\theta_t)dF(\theta_t)$ and the average treatment effect for the average student $ATE(\theta_t=\bar{\theta}_t)=E\left[\theta^{NC}_{t+1}|\bar{\theta}_t^{NC},\bar{\theta}_t^N,M_{t+1}=1\right]-E\left[\theta^{NC}_{t+1}|\bar{\theta}_t^{NC},\bar{\theta}_t^N,M_{t+1}=0\right]$ using 40,000 simulations based on the estimated parameters of the dynamic model. Standard deviation of $\theta^{NC}_{t+1}=0.623$ and of $\theta^{C}_{t+1}=1.286$.

Figure 5: Effect on Non-Cognitive $E\left[\theta^A_{t+1}|\theta_t^{NC},\theta_t^N,M_{t+1}=1\right]-E\left[\theta^A_{t+1}|\theta_t^{NC},\theta_t^N,M_{t+1}=0\right]$



Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model described in equations (1) through (4).

3.3 Allowing Non-cognitive Investment Factor Affect Cognitive Skill Production

The non-cognitive investment factor may be important in producing cognitive skills. One could consider that good parenting may affect *directly* the production of next period cognitive skills and not only *indirectly* throught its role in fostering non-cognitive skills, which consequently promote the production of cognitive skills. In this Appendix, I answer that empirical question. I estimate a model in which the non-cognitive investment factor enters the production functions of both cognitive and non-cognitive skills.

Table 11: Model of Skill Formation with Non-Cognitive Investment Only

	(1)	(2)		(3)	(4)	(5)	(6)
				$M_{t+1}=0$		$M_{t+1}=1$	
	M_{t+1}	$I_{NC,t+1}$		$\theta_{NC,t+1}$	$\theta_{C,t+1}$	$\theta_{NC,t+1}$	$\theta_{C,t+1}$
$\theta_{NC,t}$	-0.069**	0.671***	$\theta_{NC,t}$	0.950	0.166	0.907	0.094
	(0.028)	(0.074)		(0.016)	(0.022)	(0.041)	(0.058)
$\theta_{C,t}$	0.008	0.029	I_{t+1}	0.033	-0.017	0.027	0.015
	(0.009)	(0.022)		(0.012)	(0.016)	(0.029)	(0.039)
$\nabla\left(\widehat{\theta}_{NC,t}\right)$	-0.009**		ρ	-0.078	-0.069	0.368	-0.281
	(0.003)			(0.082)	(0.058)	(0.357)	(0.202)
$\nabla\left(\widehat{\theta}_{C,t}\right)$	0.005						
	(0.004)						
$\nabla\left(\text{Inc}_t\right)$	-0.004*						
	(0.002)						

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) presents the marginal effects of the estimation of equation $M_{t+1}=\mathbf{1}\left[\mathbf{X}_{t+1,M}\beta^M_{t+1}+\Lambda^M_{t+1}\Theta'_{i\in c,t}+\Lambda^{M_c}_{t+1}\nabla_{\psi,i\in c}(d)+\Gamma Z_{t+1,c}>e^M_{t+1}\right]$ of the structural model. It includes observable controls age, gender, family composition and % of troubled families (i.e., the number of peers whose families score above the mean in the violent family index). $\nabla(\cdot)$ refers to the number of classmates within a window of 10% of a SD around observation i. $\widehat{\theta}_{NC}$ is the residualized measure of self-esteem and $\widehat{\theta}_C$ is the residualized measure of the yearly test. The marginal effect of the $\nabla(\cdot)$ variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. Column (2) preset the estimates of equations $I_{A,i\in c,t+1}=\alpha^S_{A,t}\theta_{A,i\in c,t}+\alpha^S_{B,t}\theta_{B,i\in c,t}+\varepsilon_{A,i\in c,t+1}$ of the structural model. Columns (3)-(6) present the estimates of the victimization-specific production functions of non-cognitive and cognitive skills

$$\theta_{S,i\in c,t+1}=\begin{cases}\left[\gamma^0_{A,S,t}\theta^{\rho^0_S}_{A,i\in c,t}+\gamma^0_{B,S,t}\theta^{\rho^0_S}_{B,i\in c,t}+\gamma^0_{I,S,t}I^{\rho^0_S}_{A,i\in c,t+1}\right]^{1/\rho^0_S}+\eta^0_{S,i\in c,t} & \text{if } M_{i,t+1}=0\\ \left[\gamma^1_{A,S,t}\theta^{\rho^1_S}_{A,i\in c,t}+\gamma^1_{B,S,t}\theta^{\rho^1_S}_{B,i\in c,t}+\gamma^1_{I,S,t}I^{\rho^1_S}_{A,i\in c,t+1}\right]^{1/\rho^1_S}+\eta^1_{S,i\in c,t} & \text{if } M_{i,t+1}=1\end{cases}$$

for $S\in\{A,B\}$. Note that the coefficient for $\theta_{C,t}$ (i.e., $\gamma_{C,t}$) can be obtained from $\gamma_{C,t}=1-\gamma_{NC,t}-\gamma_{I,t}$.

Table 11 shows that the results remain unchanged. If anything, the share parameters of non-cognitive investment on cognitive skill development are even smaller than the investment share parameters when I use cognitive investment (Table 3 in the paper). These results indicate that, given the levels of current skills, investment—regardless of its kind—contributes very little to skill development among South Korean teenagers.

4 Heterogeneous Effects Depending on the Number of Bullies in the Classroom

The consequences of being bullied could differ depending on the degree of bullying/victimization prevailing in the classroom.[1] The KYPS-JHS collects information on bullying perpetration. Based on this information, I calculate the number of bullies in each classroom. In fact, in some specification of the treatment equation, I use that information to model the likelihood of a student being victimized. Table 5 in this Web Appendix shows that the availability of perpetrators affects positively—albeit not statistically significant at the 10% level—the chances of victimization.

In this appendix, I explore the heterogeneity in the consequences of victimization that responds to the classroom availability of bullies by allowing the model to estimate different production functions depending on the number of perpetrators in the classroom. That is, I define classrooms with high (low) concentration of bullies as those with more (less) bullies than the median classroom. Then, I develop a model in which skill production functions are allowed to differ depending on the type of classroom.

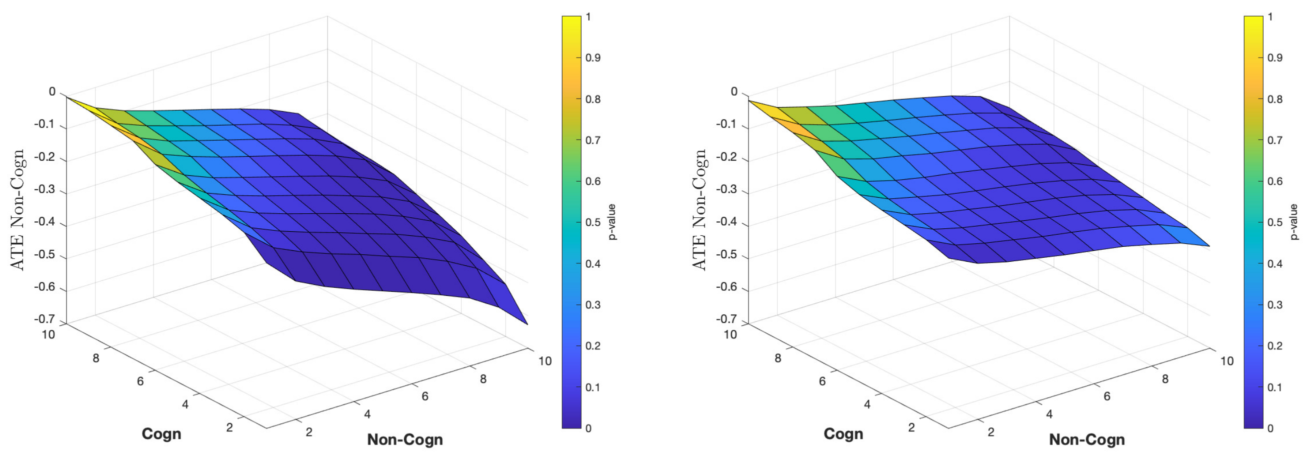
This rich model provides interesting insights. First, the negative impact on non-cognitive skill development of being bullied is larger in classrooms with lower fractions of perpetrators. In fact, bullying's overall average treatment effect on non-cognitive skills in classrooms with low concentration of bullies is -0.25 (40% of a standard deviation), while it is -0.216 (35% of a standard deviation) in classrooms with high concentration of bullies. These treatment effects are statistically different from each other. Second, when I analyze the treatment effect by skill level in Figures 6, I find that the ATE on the students in classrooms with lower fraction of bullies has a steeper gradient with respect to the initial level of non-cognitive skills that in classrooms with a high fraction of bullies. That is, victims with higher stocks of initial non-cognitive skills in classrooms with low concentration of bullies accumulate substantially less non-cognitive skills than comparable victims in classrooms with high concentration of bullies.

These results could be explained by a logic that considers that the sense of desperation might be different depending on the context where the victimization is taking place. Classroom with a higher fraction of bullies have more victims. In fact, increasing the average classroom's number of bullies by one standard deviation increases the number of victims by around 1.5 students. Another way to put it is that classrooms with above the median fraction of bullies have on average 2.5 more

[1] Thanks to an anonymous referee for this insight.

Figure 6: ATE of Being Bullying on Non-Cognitive Skill Formation

(a) Classrooms with Low Fraction of Bullies (b) Classrooms with High Fraction of Bullies



Note: Results based on 40,000 simulations based on the estimated parameters of the model of skill formation that allows for different parameters depending on the number of bullies.

victims than classrooms with below the median concentration of bullies. Thus, a victim in a high-bullying classroom has many peers that are going through the same as her, while a victim in a low-bullying classroom could feel a greater sense of desperation as she will feel she is more of a target.

References

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

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014). Fostering and Measuring Skills. OECD 110, OECD Publishing.

Sarzosa, M. and Urzua, S. (2021). Bullying among adolescents, the role of skills. *Quantitative Economics*, 12(3):945–980.

Walsh, D. (2004). *Why do they act that way?: A survival guide to the adolescent brain for you and your teen.* Free Press, New York, NY, US.